2006

Applications of nonparametric regression in survey statistics

Xiaoxi Li
Iowa State University

Applications of nonparametric regression in survey statistics

by

Xiaoxi Li

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Jean D. Opsomer, Major Professor
Song X. Chen
Barry L. Falk
Wayne A. Fuller
Michael D. Larsen

Iowa State University

Ames, Iowa

2006

UMI Number: 3243537

UMI

UMI Microform 3243537

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my family

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEGEMENTS

I would like to gratefully acknowledge my adviser Dr. Jean Opsomer's enthusiastic supervision during this work. I greatly appreciate his help for suggesting the topic and guiding me through difficulties throughout the whole time.

I also thank my PhD committee members, Dr. Wayne Fuller, Dr. Song Chen, Dr. Michael Larsen and Dr. Barry Falk for investing their time on my work and give me insightful advice. I am very thankful to other faculty and staff members at Center for Survey Statistics and Methodology (CSSM) as well for their support.

I am also grateful to all my friends at Iowa State University for being the surrogate family during the five years I stayed there.

Finally, I am forever indebted to my parents for their love and inspiration. Without them, I could not have imagine the completion of this work.

CHAPTER 1 Model-based variance estimation for systematic sampling

1.1 Systematic sampling

Systematic sampling is widely used in surveys of finite populations due to its appealing simplicity and efficiency. If applied properly, it can reflect stratifications in the population and thus can be more precise than random sampling.

Suppose that the *population size* is N and that the study variable is $Y_j \in \Re$, $j = 1, 2, \cdots, N$. Then the population mean is

$$\bar{Y}_N = \frac{1}{N}\sum_{j=1}^{N} Y_j.$$

Let n denote the *sample size* and $k = N/n$ denote the *sampling interval*. For simplicity, we assume that N is an integral multiple of n, i.e. k is an integer. Let $\mathbf{X}_j \in \Re^p$ ($j = 1, 2, \cdots, N$) be vectors of p auxiliary variables, where Y_j is some continuous and finite function of $\mathbf{X}_j$'s.

To draw a systematic sample, we first sort the population by some criterion. For example, we can sort by one of the auxiliary variables in $\mathbf{X}_j$. If the study variable Y and auxiliary variables $\mathbf{X}$ are related through a function, sorting by $\mathbf{X}$ may provide a nice 'spread' of Y's so that a systematic sample can pick up hidden stratifications in the population. If we sort the population by some criterion that is not related to Y at all, for instance, sort by a variable Z which is independent of Y, then we will have a random permutation of the population. In this case, systematic sampling is equivalent to simple random sampling without replacement (SRS).

After sorting the population, we randomly choose an element from the first k ones, say the bth one, then this systematic sample, denoted by S_b, consists of the observations with the following labels

$$b,\ b+k,\ \ldots,\ b+(n-1)k.$$

Table 1.1 illustrates this procedure. Each column corresponds to a possible systematic sample. As we can see, the interval k divides the population into n rows of k elements each. One element from each row is selected and each element has the same location on each row.

	Sample, S				
	S_1	$\cdots$	S_b	$\cdots$	S_k
Y values	Y_1	$\cdots$	Y_b	$\cdots$	Y_k
	Y_{1+k}	$\cdots$	Y_{b+k}	$\cdots$	Y_{2k}
	$\vdots$		$\vdots$		$\vdots$
	$Y_{1+(n-1)k}$	$\cdots$	$Y_{b+(n-1)k}$	$\cdots$	Y_N

Table 1.1: Systematic selection of n from population of $N = nk$ elements. Same location b is selected from n rows.

Note that there are k possible values for b, so there are k possible samples for a population. The selection probability for each element in the population is therefore $1/k$.

In systematic sampling, the population mean is estimated by the bth sample mean:

$$\bar{Y}_{S_b} = \frac{1}{n} \sum_{j \in S_b} Y_{S_b j}.$$

The design-based variance for sample mean $\bar{Y}_S$, denoted by $\mathrm{Var}_p(\bar{Y}_S)$, was first developed by Madow and Madow (1944), where

$$\mathrm{Var}_p(\bar{Y}_S) = \frac{1}{k}\sum_{b=1}^{k}(\bar{Y}_{S_b} - \bar{Y}_N)^2. \tag{1.1}$$

In $\mathrm{Var}_p(\bar{Y}_S)$, S is a member of the set of possible samples $\{S_1, \cdots, S_k\}$.

In the terminology of cluster sampling, systematic sampling is equivalent to grouping the population into k clusters, each of size n, and drawing an SRS sample of one cluster. Thus, there is no unbiased estimator for the design variance $\mathrm{Var}_p(\bar{Y}_S)$ because there are not enough degrees of freedom to estimate this quantity (Iachan, 1982).

1.2 Variance estimation for systematic sampling

As we have shown, it is impossible to derive an unbiased randomization based estimator for $\mathrm{Var}_p(\bar{Y}_S)$. Several alternative design-based approaches are discussed in the literature. One is to use biased variance estimators. Särndal et al. (1992) remarked that the estimator for SRS design, i.e.

$$\hat{\mathrm{V}}_{SRS}(\bar{Y}_S) = \frac{1-f}{n}\frac{1}{n-1}\sum_{j\in S}(Y_j - \bar{Y}_S)^2, \tag{1.2}$$

where $f = n/N$, will often overestimate the variance. A more comprehensive review can be found in Wolter (1985), where eight biased variance estimators were described and guidelines for choosing among them were given. Wolter (1985) suggested that the estimators that have good properties on average are the one based on overlapping differences (Kish 1965, sec. 4.1) , i.e.

$$\hat{\mathrm{V}}_{OL}(\bar{Y}_S) = \frac{1-f}{n}\frac{1}{2(n-1)}\sum_{j=2}^{n}(Y_j - Y_{j-1})^2, \tag{1.3}$$

and the one based on nonoverlapping differences (Kish 1965, sec. 4.1), i.e.

$$\hat{\mathrm{V}}_{NO}(\bar{Y}_S) = \frac{1-f}{n}\frac{1}{n}\sum_{j=1}^{n/2}(Y_{2j} - Y_{2j-1})^2. \tag{1.4}$$

In the case where sample size is very small, the overlapping difference estimator $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ is preferred as it has more degrees of freedom than the nonoverlapping difference estimator.

Another approach, for example, is to take more than one sample. Törnqvist (1963) discussed a method that follows the idea of interpenetrating samples introduced by Mahalanobis (1946). Let $\bar{Y}_{S_1}, \cdots, \bar{Y}_{S_c}$ be the means of *independent* systematic subsamples, then the variance of $\bar{Y}_S^* = c^{-1}\sum_{j=1}^{c}\bar{Y}_{S_j}$ is unbiasedly estimated by

$$\hat{V}(\bar{Y}_S^*) = \frac{1}{c(c-1)}\sum_{j=1}^{c}(\bar{Y}_{S_j} - \bar{Y}_S^*)^2.$$

Note that the above design gives an unbiased variance estimator, but it is essentially a different design from the systematic sampling that will be discussed in this work.

Zinger (1980) pursued an approach, defined as partially systematic sampling, that gives an unbiased and positive variance estimator, by using a systematic sample and a simple random sample from the remaining population. Let $\bar{Y}_S$ and $\bar{Y}_R$ denote the systematic sample mean and the simple random sample mean, respectively. Let n_S and n_R be the sample size of systematic sample and simple random sample, respectively. Then the combined sample mean, denoted by $\bar{Y}(\beta)$, is

$$\bar{Y}(\beta) = (1-\beta)\bar{Y}_S + \beta\bar{Y}_R,$$

with $\beta \geq 0$. The design-based variance is

$$\mathrm{Var}_p(\bar{Y}(\beta)) = b_1(\beta)S_{YU}^2 + b_2(\beta)\mathrm{Var}_p(\bar{Y}_S),$$

where

$$\begin{aligned}
S_{YU}^2 &= \frac{1}{N-1}\sum_{j\in U}(Y_i - \bar{Y}_U)^2, \\
b_1(\beta) &= \beta^2(N-1)(N-n_S-n_R)/n_R N(N-n_S-1), \\
b_2(\beta) &= (1-k\beta/(k-1))^2 - (N-n_S-n_R)\beta^2/(n_R(k-1)^2(N-n_S-1)).
\end{aligned}$$

An unbiased estimator for $\mathrm{Var}_p(\bar{Y}(\beta))$ is then

$$\hat{V}(\bar{Y}(\beta)) = b_1(\beta)s^2 + b_2(\beta)v,$$

where

$$s^2 = \frac{N}{N-1}\left[\frac{\alpha_2 Q(0) - n\alpha_1(\bar{Y}_S - \bar{Y}_R)^2}{(n_S+n_R)\alpha_2 - \alpha_1(N-n_S-n_R)/(N-n_S-1)}\right],$$

$$v = (k-1)\left[\frac{(N-n_S-n_R)Q(0)/(N-n_S-1) - n_R(n_S+n_R)(\bar{Y}_S-\bar{Y}_R)^2}{(N-n_S-n_R)\alpha_1/(N-n_S-1) - (n_S+n_R)\alpha_2}\right],$$

and

$$\begin{aligned}\alpha_1 &= n_S + n_R + kn_R - N,\\ \alpha_2 &= kn_R + n_R + [n_S(n_R-1)/(N-n_S-1)].\end{aligned}$$

The above variance estimation methods are conditional on the design. In other words, they are design-based in the sense that we treat the finite population as fixed. There also exist model-based variance estimators where the populations are considered random realizations from a superpopulation model. Wolter (1985) described a general model-based approach to estimate the variance, which is a similar approach that we will be following in this work. He pointed out that a practical difficulty with that model-base variance estimator is the correct specification of the superpopulation model. Montanari and Bartolucci (1998) proposed a linear model-based variance estimator, which is approximately unbiased for the *anticipated variance*, i.e. the expectation of the design-based variance for the sample mean under a linear superpopulation model. However, it may lack accuracy and efficiency due to a higher contribution of the bias if the systematic component of the superpopulation is significantly different from linear. A new class of unbiased estimators were proposed by Bartolucci and Montanari (2006), denoted by $\hat{\mathrm{V}}_L(\bar{Y}_S)$ in what follows. Based on that approach, they proposed two new estimators that are unbiased under linear models.

We propose a model-based nonparametric variance estimator based on local polynomial regression. Section 1.3 reviews the model-based variance results under the linear superpopulation model. In section 1.4, we study the properties of the proposed local polynomial variance estimator under the nonparametric superpopulation model. Simulation results and conclusions are presented in section 1.5.

1.3 Variance estimators under linear regression models

Due to the fact that no unbiased estimator for the design variance $\mathrm{Var}_p(\bar{Y}_S)$ exists, we consider the model-based context, where the finite population is regarded as a random realization from a superpopulation model. The following section is a continuation of Bartolucci and Montanari (2006)'s work. First, let us introduce their model-based variance estimator.

Let $Y_j \in \Re$ $(j = 1, 2, \cdots)$ be a set of independent and identically distributed random variables. Let $\mathbf{X}_j \in \Re^p$ $(j = 1, 2, \cdots)$ be vectors of auxiliary variables, which we consider to be fixed with respect to the superpopulation model.

First, let us examine the case where this superpopulation model is linear. We use L to denote the following linear model

$$\mathbf{Y} = \mathbf{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}, \tag{1.5}$$

where $\mathrm{E}_L(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{Var}_L(\boldsymbol{\varepsilon}) = \sigma_L^2 \boldsymbol{\Omega}$, with E_L and Var_L denoting the expectation and variance under the model L, respectively. We assume the errors to be mutually independent, i.e. $\boldsymbol{\Omega}$ is a diagonal matrix and $\boldsymbol{\Omega} = \mathrm{diag}\{\omega_1, \omega_2, \cdots, \omega_N\}$. Each element in $\boldsymbol{\Omega}$ is assumed to be known.

In model (1.5), $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_N)^T$, $\boldsymbol{\beta}^* = (\beta_0, \beta_1, \cdots, \beta_p)^T$, and

$$\mathbf{X}^* = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{N1} & \cdots & X_{Np} \end{pmatrix}.$$

We can write the design variance as

$$\mathrm{Var}_p(\bar{Y}_S) = \frac{1}{k}\sum_{b=1}^{k}(\bar{Y}_{S_b} - \bar{Y}_N)^2 = \frac{1}{kn^2}\mathbf{Y}^T\mathbf{D}\mathbf{Y}, \tag{1.6}$$

where $\mathbf{D} = \mathbf{M}^T\mathbf{H}\mathbf{M}$, with $\mathbf{M} = \mathbf{1}_n^T \otimes \mathbf{I}_k$ and $\mathbf{H} = \mathbf{I}_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^T$. Here $\otimes$ is the Kronecker product and $\mathbf{1}_r$ is a column vector of 1's of length r. Specifically, $\mathbf{H}$ is a $k \times k$ matrix with diagonal elements being $1 - \frac{1}{k}$ and off-diagonal element being $-\frac{1}{k}$, and $\mathbf{D}$ is a $N \times N$ matrix, which consists of n $\mathbf{H}$s in rows and n $\mathbf{H}$s in columns, i.e.

$$\mathbf{D} = \begin{pmatrix} \mathbf{H} & \mathbf{H} & \cdots & \mathbf{H} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{H} & \mathbf{H} & \cdots & \mathbf{H} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 1-\frac{1}{k} & -\frac{1}{k} & \cdots & -\frac{1}{k} \\ \vdots & \ddots & \ddots & \vdots \\ -\frac{1}{k} & -\frac{1}{k} & \cdots & 1-\frac{1}{k} \end{pmatrix}. \tag{1.7}$$

Let $\mathbf{X}$ denote the submatrix of $\mathbf{X}^*$ without the first column and $\boldsymbol{\beta}$ the subvector of $\boldsymbol{\beta}^*$ without the first element β_0. Then the *model anticipated variance* of $\bar{Y}_S$ is

$$\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] = \frac{1}{kn^2}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta} + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})\sigma_L^2. \tag{1.8}$$

Bartolucci and Montanari (2006) proposed an unbiased estimator for $\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]$, defined as

$$\hat{\mathrm{V}}_L(\bar{Y}_S) = \frac{1}{kn^2}\hat{\boldsymbol{\beta}}_b^T\mathbf{X}^T\mathbf{D}\mathbf{X}\hat{\boldsymbol{\beta}}_b - t\hat{\sigma}_{Lb}^2 + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})\hat{\sigma}_{Lb}^2. \tag{1.9}$$

where $\hat{\boldsymbol{\beta}}_b$ is the weighted least square (WLS) estimator for $\boldsymbol{\beta}$ from the bth sample and $\hat{\sigma}_{Lb}^2$ is a model unbiased estimator for σ_L^2. Specifically,

$$\hat{\boldsymbol{\beta}}_b = (\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}\mathbf{X}_b^T)^{-1}\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}\mathbf{Y}_b,$$

and

$$\hat{\sigma}^2_{Lb} = \frac{(\mathbf{Y}_b - \mathbf{X}_b\hat{\boldsymbol{\beta}}_b)^T\boldsymbol{\Omega}_b^{-1}(\mathbf{Y}_b - \mathbf{X}_b\hat{\boldsymbol{\beta}}_b)}{n - \text{rank}(\mathbf{X})},$$

where $\mathbf{X}_b$ is a submatrix of $\mathbf{X}$ and $\boldsymbol{\Omega}_b$ a submatrix of $\boldsymbol{\Omega}$ corresponding to the elements in the bth sample.

In equation (1.9), $t\hat{\sigma}^2_{Lb}$ is a bias correction term where

$$t = \frac{1}{(kn)^2}\sum_{b=1}^{k} \text{tr}(\mathbf{P}_b^T\mathbf{X}^T\mathbf{D}\mathbf{X}\mathbf{P}_b\boldsymbol{\Omega}_b),$$

and a choice of $\mathbf{P}_b$ is $\mathbf{P}_b = (\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}\mathbf{X}_b)^{-1}\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}$. We assume that $(\mathbf{X}_b^T\boldsymbol{\Omega}_b^{-1}\mathbf{X}_b)^{-1}$ exists.

We are interested in the convergence properties of $\hat{\text{V}}_L(\bar{Y}_S)$. Note that $\hat{\text{V}}_L(\bar{Y}_S)$ is an estimator of $\text{E}_L[\text{Var}_p(\bar{Y}_S)]$ and a predictor of $\text{Var}_p(\bar{Y}_S)$. We call $\hat{\text{V}}_L(\bar{Y}_S)$ a predictor of $\text{Var}_p(\bar{Y}_S)$ mainly for two reasons. One is that in this model-based context, $\text{Var}_p(\bar{Y}_S)$ is no longer a fixed value. Instead, it is a random variable associated with the random population. The other reason is that the difference between $\text{E}_L[\text{Var}_p(\bar{Y}_S)]$ and $\text{Var}_p(\bar{Y}_S)$ has small order in probability, as we will shown in Theorem 1.1.

To prove our main results in this section, we make the following assumptions.

A1.1. *The superpopulation model, denoted by L, is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where errors ε_j are mutually independent with mean zero, variance $\omega_j\sigma^2_L$.*

A1.2. *We consider all the elements in $\mathbf{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ to be fixed with respect to the superpopulation model L. We also assume that all the elements in $\mathbf{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ are bounded.*

A1.3. *The third and fourth moments of ε_j, denoted by $m_{jr} = E(\varepsilon_j)^r$, $r = 3$ and 4, exist and are bounded.*

A1.4. *Let sample size n and sampling interval k be positive integers with nk=N. We also let $n \to \infty$ and $k = O(1)$ or $k \to \infty$.*

The following theorem addresses the convergence properties of $\hat{\mathrm{V}}_L(\bar{Y}_S)$. It shows that $\hat{\mathrm{V}}_L(\bar{Y}_S)$ is a consistent estimator for $\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]$ and a consistent predictor for $\mathrm{Var}_p(\bar{Y}_S)$ and the convergence rate is $O_p(n^{-1/2})$.

Theorem 1.1. *Let $\bar{Y}_S$ denote the systematic sample mean. Let $Var_p(\bar{Y}_S)$ denote the design-based variance of $\bar{Y}_S$. Under superpopulation model L, let $E_L[Var_p(\bar{Y}_S)]$ denote the model anticipated variance of $\bar{Y}_S$ and $\hat{V}_L(\bar{Y}_S)$ an estimator for $E_L[Var_p(\bar{Y}_S)]$, where $Var_p(\bar{Y}_S)$, $E_L[Var_p(\bar{Y}_S)]$ and $\hat{V}_L(\bar{Y}_S)$ are defined as in (1.6), (1.8) and (1.9), respectively. Then, assuming A1.1 - A1.4,*

$$Var_p(\bar{Y}_S) - E_L[Var_p(\bar{Y}_S)] = O_p\left(\frac{1}{\sqrt{N}}\right), \tag{1.10}$$

$$\hat{V}_L(\bar{Y}_S) - E_L[Var_p(\bar{Y}_S)] = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{1.11}$$

$$and \qquad \hat{V}_L(\bar{Y}_S) - Var_p(\bar{Y}_S) = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{1.12}$$

Proof. First, let us prove (1.10). Denote

$$\mathbf{X}\boldsymbol{\beta} = (\sum_{i=1}^{p} X_{1i}\beta_i, \sum_{i=1}^{p} X_{2i}\beta_i, \cdots, \sum_{i=1}^{p} X_{Ni}\beta_i)^T \equiv (Q_1, Q_2, \cdots, Q_N)^T.$$

Note that $\mathrm{E}(Y_j) = Q_j$ and $\mathrm{Var}(Y_j) = \sigma_L^2\omega_j$. By assumption A1.3, for $r = 3, 4$, $m_{jr} = \mathrm{E}(Y_j - \mathrm{E}(Y_j))^r = \mathrm{E}(\varepsilon_j)^r < \infty$.

By Lemma A.1,

$$\begin{aligned}
\mathrm{Var}_L(\mathrm{Var}_p(\bar{Y}_S)) &= \mathrm{Var}_L\left(\frac{1}{kn^2}\mathbf{Y}^T\mathbf{D}\mathbf{Y}\right) \\
&= \frac{1}{k^2n^4}\mathbf{d}^T\mathbf{m}_4\mathbf{d} - \frac{3\sigma_L^4}{k^2n^4}\mathbf{d}^T\boldsymbol{\Omega}^2\mathbf{d} + \frac{2\sigma_L^4}{k^2n^4}\mathrm{tr}\left[(\mathbf{D}\boldsymbol{\Omega})^2\right] \\
&\quad + \frac{4\sigma_L^2}{k^2n^4}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\boldsymbol{\Omega}\mathbf{D}\mathbf{X}\boldsymbol{\beta} + \frac{4}{k^2n^4}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{m}_3\mathbf{d},
\end{aligned} \tag{1.13}$$

where $\mathbf{D} \equiv (d_{ij})$ is defined as in (1.7), $\mathbf{d}^T = (d_{11}, ..., d_{NN})$ and $\mathbf{m}_3$ and $\mathbf{m}_4$ are vectors of the third and fourth moment of $\mathbf{Y}$, respectively. Now let us evaluate each term on the right hand side of (1.13).

For the first term, note that the fourth moment m_{4j}'s are finite, so

$$\begin{aligned}
\frac{1}{k^2n^4}\mathbf{d}^T\mathbf{m}_4\mathbf{d} &= \frac{1}{k^2n^4}\sum_{j=1}^{N} d_{jj}^2 m_{4j} \\
&= \frac{1}{k^2n^4}\sum_{j=1}^{N}\left(1-\frac{1}{k}\right)^2 m_{4j} \\
&\le \frac{1}{k^2n^4} N\max\{m_{4j}\} \\
&= O\left(\frac{1}{Nn^2}\right). \qquad (1.14)
\end{aligned}$$

For the second term, since ω_j's are finite,

$$\begin{aligned}
\frac{3\sigma_L^4}{k^2n^4}\mathbf{d}^T\mathbf{\Omega}^2\mathbf{d} &= \frac{3\sigma_L^4}{k^2n^4}\sum_{j=1}^{N} d_{jj}^2\omega_j^2 \\
&= \frac{3\sigma_L^4}{k^2n^4}\sum_{j=1}^{N}\left(1-\frac{1}{k}\right)^2\omega_j^2 \\
&\le \frac{3\sigma_L^4}{k^2n^4} N\max\{\omega_j^2\} \\
&= O\left(\frac{1}{Nn^2}\right). \qquad (1.15)
\end{aligned}$$

For the third term, note that the jth diagonal element of $(\mathbf{D\Omega})^2$ is

$$\left(1-\frac{2}{k}\right)\omega_j\sum_{j\in S_b}\omega_j + \frac{1}{k^2}\omega_j\sum_{j\in U}\omega_j,$$

where $b = 1, 2, \cdots, k$ and S_b is the systematic sample that contains the jth observation. For instance, if $j = 1,\ k+1,\ 2k+1, \cdots, (n-1)k+1$, then $b = 1$; if $j = 2,\ k+2,\ 2k+2, \cdots, (n-1)k+2$, then $b = 2$, etc.

So

$$\begin{aligned}
\mathrm{tr}\left[(\mathbf{D\Omega})^2\right] &= \sum_{j=1}^{N}\left[\left(1-\frac{2}{k}\right)\omega_j\sum_{j\in S_b}\omega_j + \frac{1}{k^2}\omega_j\sum_{j\in U}\omega_j\right] \\
&= \left(1-\frac{2}{k}\right)\sum_{j\in U}\omega_j\sum_{j\in S_b}\omega_j + \frac{1}{k^2}\sum_{j\in U}\omega_j\sum_{j\in U}\omega_j \\
&= \left(1-\frac{2}{k}\right)\sum_{b=1}^{k}\sum_{j\in S_b}\omega_j\sum_{j\in S_b}\omega_j + \frac{1}{k^2}\sum_{j\in U}\omega_j\sum_{j\in U}\omega_j
\end{aligned}$$

$$= \left(1-\frac{2}{k}\right)\sum_{b=1}^{k}\left(\sum_{j\in S_b}\omega_j\right)^2+\frac{1}{k^2}\left(\sum_{j\in U}\omega_j\right)^2.$$

Thus,

$$\begin{aligned}
\frac{2\sigma_L^4}{k^2n^4}\mathrm{tr}\left[(\mathbf{D\Omega})^2\right] &= \frac{2\sigma_L^4}{kn^2}\left(1-\frac{2}{k}\right)\frac{1}{k}\sum_{b=1}^{k}\left(\frac{1}{n}\sum_{j\in S_b}\omega_j\right)^2+\frac{2\sigma_L^4}{k^2n^2}\left(\frac{1}{kn}\sum_{j\in U}\omega_j\right)^2\\
&= O\left(\frac{1}{kn^2}\right)+O\left(\frac{1}{k^2n^2}\right)\\
&= O\left(\frac{1}{Nn}\right). \qquad (1.16)
\end{aligned}$$

Now let us investigate the fourth term,

$$\begin{aligned}
\frac{4\sigma_L^2}{k^2n^4}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D\Omega DX}\boldsymbol{\beta} &= \frac{4\sigma_L^2}{k^2n^4}(Q_1,Q_2,\cdots,Q_N)\mathbf{D\Omega D}(Q_1,Q_2,\cdots,Q_N)^T\\
&= \frac{4\sigma_L^2}{k^2n^4}\left(T_1^2\sum_{j\in S_1}\omega_j+T_2^2\sum_{j\in S_2}\omega_j+\cdots+T_k^2\sum_{j\in S_k}\omega_j\right)\\
&= \frac{4\sigma_L^2}{k^2n^4}\left(\sum_{b=1}^{k}T_b^2\sum_{j\in S_b}\omega_j\right),
\end{aligned}$$

where $T_b=\sum_{j\in S_b}Q_j-\frac{1}{k}\sum_{j\in U}Q_j$ and $b=1,2,\cdots,k$.

Clearly, $\frac{1}{n}T_b=O(1)$ and $\frac{1}{n}\sum_{j\in S_b}\omega_j=O(1)$, so

$$\frac{4\sigma_L^2}{k^2n^4}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D\Omega DX}\boldsymbol{\beta} = \frac{4\sigma_L^2}{kn}\left(\frac{1}{k}\sum_{b=1}^{k}\left(\frac{1}{n^2}T_b^2\frac{1}{n}\sum_{j\in S_b}\omega_j\right)\right)=O\left(\frac{1}{N}\right). \qquad (1.17)$$

Finally,

$$\begin{aligned}
\frac{4}{k^2n^4}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Dm}_3\mathbf{d} &= \frac{4}{k^2n^4}(Q_1,Q_2,\cdots,Q_N)\mathbf{Dm}_3\mathbf{d}\\
&= \frac{4}{k^2n^4}\left(1-\frac{1}{k}\right)\left(T_1\sum_{j\in S_1}m_{3j}+\cdots+T_k\sum_{j\in S_k}m_{3j}\right)\\
&= \frac{4}{k^2n^4}\left(1-\frac{1}{k}\right)\left(\sum_{b=1}^{k}T_b\sum_{j\in S_b}m_{3j}\right)\\
&\le \frac{4}{kn^2}\left(\frac{1}{k}\sum_{b=1}^{k}\left(\frac{1}{n}T_b\frac{1}{n}\sum_{j\in S_b}m_{3j}\right)\right)\\
&= O\left(\frac{1}{Nn}\right). \qquad (1.18)
\end{aligned}$$

By (1.14), (1.15), (1.16), (1.17) and (1.18),

$$\mathrm{Var}_L(\mathrm{Var}_p(\bar{Y}_S)) = O\left(\frac{1}{N}\right). \tag{1.19}$$

Note that

$$\mathrm{Var}_L(\mathrm{Var}_p(\bar{Y}_S)) = \mathrm{E}\left(\mathrm{Var}_p(\bar{Y}_S) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)]\right)^2,$$

thus by Corollary A.1,

$$\mathrm{Var}_p(\bar{Y}_S) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{1.20}$$

and thus (1.10) holds.

Secondly, let us prove (1.11). Note that

$$\begin{aligned}
& \hat{\mathrm{V}}_L(\bar{Y}_S) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] \\
= & \frac{1}{kn^2}\hat{\boldsymbol{\beta}}_b^T \mathbf{X}^T\mathbf{D}\mathbf{X}\hat{\boldsymbol{\beta}}_b - t\hat{\sigma}_{Lb}^2 + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})\hat{\sigma}_{Lb}^2 \\
& -\frac{1}{kn^2}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta} - \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})\sigma_L^2 \\
= & \frac{1}{kn^2}\Big[(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^T\mathbf{X}^T\mathbf{D}\mathbf{X}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) \\
& +(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta}\Big] + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})(\hat{\sigma}_{Lb}^2 - \sigma_L^2) - t\hat{\sigma}_{Lb}^2.
\end{aligned} \tag{1.21}$$

We will examine each term on the right hand side of (1.21).

For linear regression models, $\hat{\beta}_{bi} - \beta_i = O_p\left(\frac{1}{\sqrt{n}}\right)$ and $\hat{\sigma}_{Lb}^2 - \sigma_L^2 = O_p\left(\frac{1}{\sqrt{n}}\right)$. The element on ith row and jth column of $\frac{1}{kn^2}\mathbf{X}^T\mathbf{D}\mathbf{X}$ is

$$\frac{1}{k}\sum_{b=1}^{k}(\bar{X}_{S_b i} - \bar{X}_{Ui})(\bar{X}_{S_b j} - \bar{X}_{Uj}) = O(1), \tag{1.22}$$

where $i, j = 1, ..., p$. So

$$\frac{1}{kn^2}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^T\mathbf{X}^T\mathbf{D}\mathbf{X}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) = O_p\left(\frac{1}{n}\right), \tag{1.23}$$

and

$$\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta} = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{1.24}$$

Also note that

$$\begin{aligned}
\frac{1}{kn^2}\mathrm{tr}(\mathbf{D\Omega}) &= \frac{1}{kn^2}\mathrm{tr}\left[\begin{pmatrix} \mathbf{H} & \cdots & \mathbf{H} \\ \vdots & \ddots & \vdots \\ \mathbf{H} & \cdots & \mathbf{H} \end{pmatrix}\begin{pmatrix} \omega_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_N \end{pmatrix}\right] \\
&= \frac{1}{kn^2}\left(1-\frac{1}{k}\right)\sum_{j=1}^{N}\omega_j \\
&= \frac{1}{n}\left(1-\frac{1}{k}\right)\frac{1}{N}\sum_{j=1}^{N}\omega_j \\
&= O\left(\frac{1}{n}\right). \qquad (1.25)
\end{aligned}$$

Thus,

$$\frac{1}{kn^2}\mathrm{tr}(\mathbf{D\Omega})(\hat{\sigma}^2_{Lb}-\sigma^2_L) = O_p\left(\frac{1}{n\sqrt{n}}\right). \qquad (1.26)$$

Now it remains to show the order in probability of $t\hat{\sigma}^2_{Lb}$. By the definition of t,

$$\begin{aligned}
t &= \frac{1}{(kn)^2}\sum_{b=1}^{k}\mathrm{tr}(\mathbf{P}_b^T\mathbf{X}^T\mathbf{D}\mathbf{X}\mathbf{P}_b\mathbf{\Omega}_b) \\
&= \frac{1}{(kn)^2}\sum_{b=1}^{k}\mathrm{tr}(\underbrace{\mathbf{\Omega}_b^{-1}\mathbf{X}_b}_{A_1}\underbrace{(\mathbf{X}_b^T\mathbf{\Omega}_b^{-1}\mathbf{X}_b)^{-1}\mathbf{X}^T\mathbf{D}\mathbf{X}(\mathbf{X}_b^T\mathbf{\Omega}_b^{-1}\mathbf{X}_b)^{-1}\mathbf{X}_b^T}_{A_2}) \\
&= \frac{1}{(kn)^2}\sum_{b=1}^{k}\mathrm{tr}(\underbrace{(\mathbf{X}_b^T\mathbf{\Omega}_b^{-1}\mathbf{X}_b)^{-1}\mathbf{X}^T\mathbf{D}\mathbf{X}(\mathbf{X}_b^T\mathbf{\Omega}_b^{-1}\mathbf{X}_b)^{-1}\mathbf{X}_b^T}_{A_2}\underbrace{\mathbf{\Omega}_b^{-1}\mathbf{X}_b}_{A_1}) \\
&= \frac{1}{(kn)^2}\sum_{b=1}^{k}\mathrm{tr}((\mathbf{X}_b^T\mathbf{\Omega}_b^{-1}\mathbf{X}_b)^{-1}\mathbf{X}^T\mathbf{D}\mathbf{X}).
\end{aligned}$$

Let $\mathbf{\Omega}_b$ denote the variance-covariance matrix for observations from the bth sample, then $\boldsymbol{\Omega}_b = \text{diag}\{\omega_b, \omega_{b+k}, \cdots, \omega_{b+(n-1)k}\}$, and

$$\mathbf{X}_b = \begin{pmatrix} X_{b1} & \cdots & X_{bp} \\ \vdots & \ddots & \vdots \\ X_{(b+(n-1)k)1} & \cdots & X_{(b+(n-1)k)p} \end{pmatrix}.$$

The element on the ith row and jth column of $\frac{1}{n}\mathbf{X}_b^T \boldsymbol{\Omega}_b^{-1} \mathbf{X}_b$ is

$$\frac{1}{n} \sum_{t=0}^{n-1} \frac{X_{(b+tk)i} X_{(b+tk)j}}{\omega_{b+tk}} = O(1), \tag{1.27}$$

where $i, j = 1, ..., p$.

Note that $\text{tr}((\frac{1}{n}\mathbf{X}_b^T \boldsymbol{\Omega}_b^{-1} \mathbf{X}_b)^{-1}(\frac{1}{kn^2}\mathbf{X}^T \mathbf{D} \mathbf{X}))$ is a sum of p terms. By (1.22) and (1.27), each of those p terms is finite, so

$$\text{tr}((\frac{1}{n}\mathbf{X}_b^T \boldsymbol{\Omega}_b^{-1} \mathbf{X}_b)^{-1}(\frac{1}{kn^2}\mathbf{X}^T \mathbf{D} \mathbf{X})) = O(1).$$

Thus

$$\begin{aligned} t &= \frac{1}{(kn)^2} \sum_{b=1}^{k} \frac{kn^2}{n} \text{tr}\left(\left(\frac{1}{n}\mathbf{X}_b^T \boldsymbol{\Omega}_b^{-1} \mathbf{X}_b \right)^{-1} \left(\frac{1}{kn^2}\mathbf{X}^T \mathbf{D} \mathbf{X} \right) \right) \\ &= O\left(\frac{1}{n}\right). \end{aligned} \tag{1.28}$$

Also note that $\hat{\sigma}_{Lb}^2 = \sigma_L^2 + O_p\left(\frac{1}{\sqrt{n}}\right) = O(1) + O_p\left(\frac{1}{\sqrt{n}}\right) = O_p(1)$. Therefore,

$$t\hat{\sigma}_{Lb}^2 = O_p\left(\frac{1}{n}\right). \tag{1.29}$$

By (1.23), (1.24), (1.26) and (1.29), equation (1.11) holds.

Finally, (1.12) follows from (1.10) and (1.11). □

Remark 1: Although we did not calculate the order of $\text{E}_L[\text{Var}_p(\bar{Y}_S)]$ in the proof of Theorem 1.1, it is useful to describe it here. Note that $\text{E}_L\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$, $\text{Var}_L\mathbf{Y} = \boldsymbol{\Omega}\sigma_L^2$ and

$\mathbf{D}$ is a symmetric matrix defined in (1.7). By Theorem A.1,

$$\begin{aligned}\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] &= \mathrm{E}_L[\frac{1}{kn^2}\mathbf{Y}^T\mathbf{D}\mathbf{Y}] \\ &= \frac{1}{kn^2}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta} + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})\sigma_L^2. \end{aligned} \tag{1.30}$$

Let $\mathbf{X}\boldsymbol{\beta} = (\sum_{i=1}^p X_{1i}\beta_i, \sum_{i=1}^p X_{2i}\beta_i, \cdots, \sum_{i=1}^p X_{Ni}\beta_i)^T \equiv (Q_1, Q_2, \cdots, Q_N)^T$, then

$$\begin{aligned}\frac{1}{kn^2}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta} &= \frac{1}{kn^2}(Q_1, Q_2, \cdots, Q_N)\mathbf{D}(Q_1, Q_2, \cdots, Q_N)^T \\ &= \frac{1}{k}\sum_{b=1}^{k}\left(\frac{1}{n}\sum_{j\in S_b} Q_j - \frac{1}{N}\sum_{j\in U} Q_j\right)^2 \\ &= O(1). \end{aligned} \tag{1.31}$$

By (1.25),

$$\frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})\sigma_L^2 = O\left(\frac{1}{n}\right).$$

So

$$\mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] = O(1). \tag{1.32}$$

Equation (1.32) suggests that for systematic sampling, the design-based variance of the sample mean is bounded and does not generally decrease as the sample size increases.

Remark 2: Equation (1.32) implies that $\mathrm{Var}_p(\bar{Y}_S) = O_p(1)$ with respect to model L. This is because $\mathrm{Var}_p(\bar{Y}_S) - \mathrm{E}_L[\mathrm{Var}_p(\bar{Y}_S)] = O_p\left(\frac{1}{\sqrt{N}}\right)$.

Remark 3: In equation (1.31), the order of $\frac{1}{kn^2}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta}$ is actually $O\left(\left(1-\frac{1}{k}\right)^2\right)$ because

$$\begin{aligned}\frac{1}{n}\sum_{j\in S_b} Q_j - \frac{1}{N}\sum_{j\in U} Q_j &= \frac{1}{n}\sum_{j\in S_b} Q_j - \frac{1}{N}\sum_{j\in S_b} Q_j - \frac{1}{N}\sum_{j\in U\setminus S_b} Q_j \\ &= \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{j\in S_b} Q_j - \frac{1}{N}\sum_{j\in U\setminus S_b} Q_j \\ &= \left(\frac{1}{n} - \frac{1}{N}\right) n\bar{Q}_{S_b} - \frac{1}{N}(N-n)\bar{Q}_{U\setminus S_b}\end{aligned}$$

$$
\begin{aligned}
&= \left(1-\frac{n}{N}\right)\left(\bar{Q}_{S_b}-\bar{Q}_{U\setminus S_b}\right)\\
&= O\left(1-\frac{n}{N}\right)\\
&= O\left(1-\frac{1}{k}\right),
\end{aligned}
$$

which implies that

$$\frac{1}{kn^2}\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\boldsymbol{\beta} = O\left(\left(1-\frac{1}{k}\right)^2\right).$$

Note that this is essentially $O(1)$ because in systematic sampling, k is either a constant or an increasing number. Either way, $O\left(\left(1-\frac{1}{k}\right)^2\right) = O(1)$. However, this term could be of a smaller order if X's are random variables and $\bar{Q}_{S_b}$ and $\bar{Q}_U$ have the same distribution, in which case $\bar{Q}_{S_b}$ estimates $\bar{Q}_U$ extremely well. For example, when the population is sorted by model variable X, the difference between $\bar{Q}_{S_b}$ and $\bar{Q}_U$ depends on the size of sampling interval k, which is also a function of n. So as n increases, $|\bar{Q}_{S_b}-\bar{Q}_U|$ will decrease towards zero.

1.4 Variance estimators under nonparametric models

A parametric method is efficient when we correctly specify the superpopulation model. However, if the superpopulation model is incorrectly specified, a parametric method may result in biased or inefficient estimation.

In this section, we propose a consistent variance estimator under a nonparametric model. We study the case where $p=1$, i.e. one predictor variable X. Let $\mathbf{X} = (X_1, X_2, \cdots, X_N)$. Our nonparametric superpopulation model, denoted by NP, is

$$\mathbf{Y} = \mathbf{m} + \boldsymbol{\varepsilon}, \tag{1.33}$$

where $\mathrm{E}_{NP}(Y_j|X_j=x_j) = m(x_j)$ and $\mathrm{Var}_{NP}(\boldsymbol{\varepsilon}) = \sigma^2_{NP}\mathrm{diag}\{\omega_1,\omega_2,\cdots,\omega_N\} \equiv \sigma^2_{NP}\boldsymbol{\Omega}$. Let $m(\cdot)$ be a continuous and bounded function, and define $\mathbf{m} = (m(x_1),\cdots,m(x_N))$.

Then $\mathbf{m}$ is a vector of bounded numbers. We assume that the ω_j's are bounded and positive, where $j = 1, ..., N$.

As shown in (1.6), $\mathrm{Var}_p(\bar{Y}_S) = \frac{1}{k}\sum_{b=1}^{k}(\bar{Y}_{S_b} - \bar{Y}_N)^2 = \frac{1}{kn^2}\mathbf{Y}^T\mathbf{D}\mathbf{Y}$, so by Theorem A.1, the model anticipated variance of $\bar{Y}_S$ under model NP, is

$$\mathrm{E}_{NP}[\mathrm{Var}_p(\bar{Y}_S)] = \frac{1}{kn^2}\mathbf{m}^T\mathbf{D}\mathbf{m} + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\mathbf{\Omega})\sigma_{NP}^2. \tag{1.34}$$

To estimate $\mathrm{E}_{NP}[\mathrm{Var}_p(\bar{Y}_S)]$, we propose the following estimator

$$\hat{\mathrm{V}}_{NP}(\bar{Y}_S) = \frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{D}\hat{\mathbf{m}}_b) + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\mathbf{\Omega})\hat{\sigma}_{NPb}^2, \tag{1.35}$$

where $\hat{\sigma}_{NPb}^2$ is defined as

$$\hat{\sigma}_{NPb}^2 = \frac{(\mathbf{Y}_b - \hat{\mathbf{m}}_b)^T\mathbf{\Omega}_b^{-1}(\mathbf{Y}_b - \hat{\mathbf{m}}_b)}{n}, \tag{1.36}$$

and $\hat{\mathbf{m}}_b = (\hat{m}(x_1), \cdots, \hat{m}(x_n))$, where $\hat{m}(x_j)$ is the local polynomial regression estimator obtained from bth sample:

$$\hat{m}(x_j) = \mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{Y}_b.$$

Let q be the degree of local polynomial regression. We only consider the case where q is *odd*. The most popular examples are local linear regression and local cubic regression. Then $\mathbf{e}_1$ is the $(q+1)\times 1$ vector having 1 in the first entry and all other entries 0, and

$$\mathbf{X}_{bj} = \begin{pmatrix} 1 & (x_1 - x_j) & \cdots & (x_1 - x_j)^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_n - x_j) & \cdots & (x_n - x_j)^q \end{pmatrix},$$

$$\mathbf{W}_{bj} = \mathrm{diag}\left\{K\left(\frac{x_i - x_j}{h}\right), \quad i = 1, \cdots n\right\},$$

where h is the bandwidth, and K is the kernel function. For simplicity, we will use K_{ij} to denote $K\left(\frac{x_i - x_j}{h}\right)$ in future notation. Please refer to Wand and Jones (1995) for more information on local polynomial regression. Note that we are not including the bias correction term $t\hat{\sigma}^2_{Lb}$ used in equation (1.9) because that term is asymptotically negligible.

To prove our main results in this section, we need assumption A1.3 and A1.4 from the previous section. In addition, we make the following assumptions.

A1.5. *The superpopulation model, denoted by NP, is* $\mathbf{Y} = \mathbf{m} + \boldsymbol{\varepsilon}$*, where errors* ε_j *are independent with mean zero, variance* $\omega_j \sigma^2_{NP}$*.*

A1.6. *We consider* x_j*'s as fixed with respect to the superpopulation model* NP*. The* x_j*'s are independent and identically distributed with* $F(x) = \int_{-\infty}^{x} f(t)dt$*, where* $f(\cdot)$ *is a density function with compact support* $[a_x, b_x]$ *and* $f(x) > 0$ *for all* $x \in [a_x, b_x]$*. In addition, we assume that the first derivative of* f *exists.*

A1.7. *As* $n \to \infty$*,* $h \to 0$ *and* $nh \to \infty$*.*

A1.8. *The kernel function* $K(\cdot)$ *is a compactly supported, bounded kernel such that* $\int u^{q+1} K(u)du = \mu_{q+1}(K)$*, where* $\mu_{q+1}(K) \neq 0$*. In addition, all odd-order moments of* K *vanish, that is,* $\int u^q K(u)du = 0$*.*

A1.9. *The* $(q+1)th$ *derivative of the model function* $m(\cdot)$ *exists and is bounded on* $[a_x, b_x]$*.*

The following theorem addresses the convergence properties of $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$. It suggests that $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is a consistent estimator for $\mathrm{E}_{NP}[\mathrm{Var}_p(\bar{Y}_S)]$ and a consistent predictor for $\mathrm{Var}_p(\bar{Y}_S)$.

Theorem 1.2. *Let* $\bar{Y}_S$ *denote the systematic sample mean. Let* $Var_p(\bar{Y}_S)$ *denote the design-based variance of* $\bar{Y}_S$*. Under superpopulation model NP, let* $E_{NP}[Var_p(\bar{Y}_S)]$ *denote the model anticipated variance of* $\bar{Y}_S$ *and* $\hat{V}_{NP}(\bar{Y}_S)$ *our proposed estimator for*

$E_{NP}[Var_p(\bar{Y}_S)]$, where $Var_p(\bar{Y}_S)$, $E_{NP}[Var_p(\bar{Y}_S)]$ and $\hat{V}_{NP}(\bar{Y}_S)$ are defined as in (1.6), (1.34) and (1.35), respectively. Then, assuming A1.3, A1.4, and A1.5 - A1.9,

$$Var_p(\bar{Y}_S) - E_{NP}[Var_p(\bar{Y}_S)] = O_p\left(\frac{1}{\sqrt{N}}\right), \tag{1.37}$$

$$\hat{V}_{NP}(\bar{Y}_S) - E_{NP}[Var_p(\bar{Y}_S)] = O_p(h^{q+1}) + O_p\left(\frac{1}{\sqrt{nh}}\right), \tag{1.38}$$

and

$$\hat{V}_{NP}(\bar{Y}_S) - Var_p(\bar{Y}_S) = O_p(h^{q+1}) + O_p\left(\frac{1}{\sqrt{nh}}\right). \tag{1.39}$$

Note that the nonparametric variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is still a consistent estimator for the model anticipated variance, but the convergence rate is not as fast as that of the linear variance estimator $\hat{\mathrm{V}}_L(\bar{Y}_S)$, which is $\frac{1}{\sqrt{n}}$. Furthermore, the best bandwidth h should satisfy this condition: $h^{q+1} = O\left(\frac{1}{\sqrt{nh}}\right)$, which leads to $h = cn^{-1/(2q+3)}$.

Proof. First, equation (1.37) can be proved similarly to (1.10) in Theorem 1.1. Secondly, note that if (1.38) holds, then (1.39) follows by (1.37) and (1.38). So the problem remains to prove (1.38).

Note that

$$\begin{aligned}
&\hat{\mathrm{V}}_{NP}(\bar{Y}_S) - \mathrm{E}_{NP}[\mathrm{Var}_p(\bar{Y}_S)] \\
&\quad = \frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{D}\hat{\mathbf{m}}_b - \mathbf{m}^T\mathbf{D}\mathbf{m}) + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\boldsymbol{\Omega})(\hat{\sigma}^2_{NPb} - \sigma^2_{NP}).
\end{aligned} \tag{1.40}$$

The first term on the right hand side of (1.40) can be written as

$$\begin{aligned}
\frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{D}\hat{\mathbf{m}}_b - \mathbf{m}^T\mathbf{D}\mathbf{m}) &= \frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) + \frac{1}{kn^2}\mathbf{m}^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) \\
&\quad + \frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}\mathbf{m} \\
&= \frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) + \frac{2}{kn^2}\mathbf{m}^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) \\
&\equiv (A) + 2(B).
\end{aligned}$$

Note that $\mathbf{m}^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) = (\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}\mathbf{m}$ because they are both scalars. By the definition of matrix $\mathbf{D}$, we can write (A) as

$$
\begin{aligned}
(A) &= \frac{1}{k}\sum_{b=1}^{k}\left[\frac{1}{n}\sum_{j\in S_b}(\hat{m}(x_j)-m(x_j)) - \frac{1}{N}\sum_{j\in U}(\hat{m}(x_j)-m(x_j))\right]^2 \\
&= \frac{1}{k}\sum_{b=1}^{k}\left\{\frac{1}{n^2}\left[\sum_{j\in S_b}(\hat{m}(x_j)-m(x_j))\right]^2 + \frac{1}{N^2}\left[\sum_{j\in U}(\hat{m}(x_j)-m(x_j))\right]^2\right. \\
&\quad \left. -\frac{2}{nN}\sum_{j\in S_b}(\hat{m}(x_j)-m(x_j))\sum_{j\in U}(\hat{m}(x_j)-m(x_j))\right\} \\
&\equiv \frac{1}{k}\sum_{b=1}^{k}\{(a1)+(a2)+(a3)\},
\end{aligned}
$$

where

$$
\begin{aligned}
(a1) &= \frac{1}{n^2}[\sum_{j\in S_b}(\hat{m}(x_j)-m(x_j))]^2, \\
(a2) &= \frac{1}{N^2}[\sum_{j\in U}(\hat{m}(x_j)-m(x_j))]^2, \\
\text{and}\quad (a3) &= -\frac{2}{nN}\sum_{j\in S_b}(\hat{m}(x_j)-m(x_j))\sum_{j\in U}(\hat{m}(x_j)-m(x_j)).
\end{aligned}
$$

Below is the outline for the proof of (1.38).

1. Evaluate $(a1)$, $(a2)$ and $(a3)$ to get the order in probability for (A).

2. Evaluate (B) and combine the result for (A) to get the order in probability for $\frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{D}\hat{\mathbf{m}}_b - \mathbf{m}^T\mathbf{D}\mathbf{m})$.

3. Evaluate $\frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\mathbf{\Omega})(\hat{\sigma}^2_{NPb} - \sigma^2_{NP})$.

Now let us expand the parentheses in $(a1), (a2)$ and $(a3)$, then we have

$$
\begin{aligned}
(a1) &= \frac{1}{n^2}\sum_{j\in S_b}(\hat{m}(x_j)-m(x_j))^2 \\
&\quad +\frac{1}{n^2}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}(\hat{m}(x_j)-m(x_j))(\hat{m}(x_l)-m(x_l)),
\end{aligned} \tag{1.41}
$$

$$
\begin{aligned}
(a2) &= \frac{1}{N^2}\sum_{j\in U}\left(\hat{m}(x_j)-m(x_j)\right)^2 \\
&\quad +\frac{1}{N^2}\sum_{j\in U}\sum_{l\in U, j\neq l}\left(\hat{m}(x_j)-m(x_j)\right)\left(\hat{m}(x_l)-m(x_l)\right), \\
(a3) &= -\frac{2}{nN}\sum_{j\in S_b}\sum_{l\in U}\left(\hat{m}(x_j)-m(x_j)\right)\left(\hat{m}(x_l)-m(x_l)\right).
\end{aligned}
$$

Let $\mathbf{s}_{bj}$ denote the smoother matrix for local polynomial regression, where $\mathbf{s}_{bj} = \mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}$. So

$$
\hat{m}(x_j)-m(x_j)=\mathbf{s}_{bj}\mathbf{Y}_b-m(x_j)=\mathbf{s}_{bj}(\mathbf{m}_b+\boldsymbol{\varepsilon}_b)-m(x_j)=b_b(x_j)+\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b, \qquad (1.42)
$$

where $b_b(x_j)$ is the bias of local polynomial regression fitting.

Now,

$$
\begin{aligned}
\mathrm{E}(a1) &= \frac{1}{n^2}\sum_{j\in S_b}b_b^2(x_j)+\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right)+\frac{1}{n^2}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}b_b(x_j)b_b(x_l) \\
&\quad +\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right). \qquad (1.43)
\end{aligned}
$$

The right-hand side of (1.43) contains four terms. We will calculate them one by one.

(i) First let us evaluate $\frac{1}{n^2}\sum_{j\in S_b}b_b^2(x_j)$. We use a technique similar to Ruppert and Wand (1994). By Taylor's theorem,

$$
\begin{aligned}
m(x_i) &= m(x_j)+m'(x_j)(x_i-x_j)+\frac{1}{2}m''(x_j)(x_i-x_j)^2+ \\
&\quad \cdots+\frac{1}{q!}m^{(q)}(x_j)(x_i-x_j)^q+\frac{1}{(q+1)!}m^{(q+1)}(x_{ji}^*)(x_i-x_j)^{q+1},
\end{aligned}
$$

where x_{ji}^* is some point between x_i and x_j.

Let $\mathbf{m}_b=(m(x_1),m(x_2),\cdots,m(x_n))$. Then the matrix form of Taylor's expansion for $\mathbf{m}_b$ is

$$
\mathbf{m}_b=\mathbf{X}_{bj}\begin{pmatrix} m(x_j) \\ \\ \mathbf{D}_m(x_j)\end{pmatrix}+\mathbf{R}_m(x_j), \qquad (1.44)
$$

where $\mathbf{D}_m(x_j) = (m'(x_j), \frac{1}{2}m''(x_j), \cdots, \frac{1}{q!}m^{(q)}(x_j))$ and $\mathbf{R}_m(x_j)$ is a vector of Taylor series remainder terms. So

$$\begin{aligned} b_b(x_j) &= \mathbf{s}_{bj}\mathbf{m}_b - m(x_j) = \mathbf{s}_{bj}\mathbf{R}_m(x_j) \qquad (1.45) \\ &= \mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj} \begin{pmatrix} \frac{1}{(q+1)!}m^{(q+1)}(x_{j1}^*)(x_1 - x_j)^{q+1} \\ \vdots \\ \frac{1}{(q+1)!}m^{(q+1)}(x_{jn}^*)(x_n - x_j)^{q+1} \end{pmatrix} \end{aligned}$$

Let $\mu_t = \int u^t K(u)du$. Let $\mathbf{N}_q$ be the $(q+1)\times(q+1)$ matrix having (s,t)th entry equal to μ_{s+t-2} and let $\mathbf{Q}_q$ be the $(q+1)\times(q+1)$ matrix having (s,t)th entry equal to μ_{s+t-1}. Let $\mathbf{A} = \text{diag}\{1, h, ..., h^q\}$. Similar to what was done in the proof of Theorem 4.1 in Ruppert and Wand (1994), except that we consider x_j's fixed instead of random,

$$n^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj} = \mathbf{A}\{f(x_j)\mathbf{N}_q + hf'(x_j)\mathbf{Q}_q\}\mathbf{A} + o(h\mathbf{A}\mathbf{1}\mathbf{A}),$$

which leads to

$$\begin{aligned} &\mathbf{e}_1^T(n^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1} \\ = \;& f(x_j)^{-1}\{\mathbf{e}_1^T\mathbf{N}_q^{-1} - hf'(x_j)f(x_j)^{-1}\mathbf{e}_1^T\mathbf{N}_q^{-1}\mathbf{Q}_q\mathbf{N}_q^{-1}\}\mathbf{A}^{-1} + o(h\mathbf{1}\mathbf{A}^{-1}). \qquad (1.46) \end{aligned}$$

Note that on the right hand side of (1.46), the leading term is $f(x_j)^{-1}\mathbf{e}_1^T\mathbf{N}_q^{-1}\mathbf{A}^{-1}$. This is because as $h \to 0$, the term $hf'(x_j)f(x_j)^{-1}\mathbf{e}_1^T\mathbf{N}_q^{-1}\mathbf{Q}_q\mathbf{N}_q^{-1}\mathbf{A}^{-1}$ diminishes.

For $r = 0, 1, ...,$ standard results from kernel density estimation lead to

$$\mathbf{A}^{-1}n^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj} \begin{pmatrix} (x_1 - x_j)^r \\ \vdots \\ (x_n - x_j)^r \end{pmatrix}$$

$$= h^r f(x_j)\begin{pmatrix}\mu_r\\ \vdots\\ \mu_{r+q}\end{pmatrix} + h^{r+1}f'(x_j)\begin{pmatrix}\mu_{r+1}\\ \vdots\\ \mu_{r+q+1}\end{pmatrix} + o(h^{r+1}). \quad (1.47)$$

Combining (1.46) and (1.47), we obtain

$$\begin{aligned} b_b(x_j) &= \left(\sum_{t=1}^{q+1}(\mathbf{N}_q^{-1})_{1t}\mu_{q+t}\right)\frac{m^{(q+1)}(x_{ji}^*)}{(q+1)!}h^{q+1} \\ &\quad + \left(\sum_{t=1}^{q+1}(\mathbf{N}_q^{-1})_{1t}\mu_{q+t+1} - \mathbf{e}_1^T\mathbf{N}_q^{-1}\mathbf{Q}_q\mathbf{N}_q^{-1}(\mu_{q+1},\cdots,\mu_{2q+1})^T\right) \\ &\quad \cdot\frac{m^{(q+1)}(x_{ji}^*)f'(x_j)}{f(x_j)(q+1)!}h^{q+2} + o(h^{p+2}). \end{aligned} \quad (1.48)$$

We will only study the case where q is odd, and note that (a) $\mu_t = 0$ for t odd, (b) $(\mathbf{N}_q)_{st} = (\mathbf{N}_q^{-1})_{st} = 0$ for $s+t$ odd and (c) $(\mathbf{Q}_q)_{st} = 0$ for $s+t$ even. Hence, the first term on the right hand side of (1.48) does not vanish due to (a) and (b), and this is the leading bias term. So

$$b_b(x_j) = O(h^{q+1}), \quad (1.49)$$

and thus

$$\frac{1}{n^2}\sum_{j\in S_b} b^2(x_j) = \frac{1}{n^2}\sum_{j\in S_b} O(h^{2q+2}) = O\left(\frac{h^{2q+2}}{n}\right). \quad (1.50)$$

(ii) Secondly, Let us compute $\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right)$ in (1.43). Note that

$$\begin{aligned} \frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right) &= \frac{\sigma_{NP}^2}{n^2}\sum_{j\in S_b}\mathbf{s}_{bj}\boldsymbol{\Omega}_b\mathbf{s}_{bj}^T \\ = \frac{\sigma_{NP}^2}{n^2}\sum_{j\in S_b}&\mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bj}^T\mathbf{X}_{bj}(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}, \end{aligned}$$

where $\boldsymbol{\Omega}_b$ is a sub-matrix of $\boldsymbol{\Omega}$ corresponding to the bth sample. Let $\left(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bj}^T\mathbf{X}_{bj}\right)_{st}$ denote the (s,t)th entry of $\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bj}^T\mathbf{X}_{bj}$. Note that

$$\left|\left(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bj}^T\mathbf{X}_{bj}\right)_{st}\right| = \left|\sum_{i\in S_b}K_{ij}^2\omega_i(x_i-x_j)^{s+t-2}\right|$$

$$\leq \max\{\omega_i\} \sum_{i \in S_b} K_{ij}^2 |x_i - x_j|^{s+t-2} = \max\{\omega_i\} \left| \left(\mathbf{X}_{bj}^{*T} \mathbf{W}_{bj}^2 \mathbf{X}_{bj}^* \right)_{st} \right|,$$

where $\mathbf{X}_{bj}^*$'s elements are the absolute values of matrix $\mathbf{X}_{bj}$, i.e.

$$\mathbf{X}_{bj}^* = \begin{pmatrix} 1 & |x_1 - x_j| & \cdots & |x_1 - x_j|^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & |x_n - x_j| & \cdots & |x_n - x_j|^q \end{pmatrix}.$$

So the order of $\mathbf{X}_{bj}^T \mathbf{W}_{bj} \mathbf{\Omega}_b \mathbf{W}_{bj}^T \mathbf{X}_{bj}$ is the same as the order of $\mathbf{X}_{bj}^{*T} \mathbf{W}_{bj}^2 \mathbf{X}_{bj}^*$. Similar to Ruppert and Wand (1994), p.1365,

$$n^{-1} \mathbf{X}_{bj}^{*T} \mathbf{W}_{bj}^2 \mathbf{X}_{bj}^* = h^{-1} f(|x_j|) \mathbf{A} \mathbf{T}_q \mathbf{A} + o(h^{-1} \mathbf{A} \mathbf{1} \mathbf{A}),$$

where $\mathbf{T}_q$ is the $(q+1) \times (q+1)$ matrix whose (s,t)th entry is $\int u^{s+t-2} K(u)^2 du$. Combine this with (1.46), and ignore the terms of small orders, we have

$$\begin{aligned} & \frac{\sigma_{NP}^2}{n^2} \sum_{j \in S_b} \mathbf{e}_1^T (\mathbf{X}_{bj}^T \mathbf{W}_{bj} \mathbf{X}_{bj})^{-1} \mathbf{X}_{bj}^{*T} \mathbf{W}_{bj}^2 \mathbf{X}_{bj}^* (\mathbf{X}_{bj}^T \mathbf{W}_{bj} \mathbf{X}_{bj})^{-1} \\ \doteq\; & \frac{\sigma_{NP}^2}{n^3} \sum_{j \in S_b} f(x_j)^{-1} \mathbf{e}_1^T \mathbf{N}_q^{-1} \mathbf{A}^{-1} h^{-1} f(|x_j|) \mathbf{A} \mathbf{T}_q \mathbf{A} f(x_j)^{-1} \mathbf{A}^{-1} \mathbf{N}_q^{-1} \mathbf{e}_1 \\ =\; & \frac{\sigma_{NP}^2}{n^3} \sum_{j \in S_b} h^{-1} f(x_j)^{-2} f(|x_j|) \mathbf{e}_1^T \mathbf{N}_q^{-1} \mathbf{T}_q \mathbf{N}_q^{-1} \mathbf{e}_1, \end{aligned}$$

where the leading term $\mathbf{e}_1^T \mathbf{N}_q^{-1} \mathbf{T}_q \mathbf{N}_q^{-1} \mathbf{e}_1$ is of order O(1). The reasoning is as follows.

Let c_{st} denote the (s,t)th element of the cofactor matrix of $(\mathbf{N}_q)_{st}$. Note that $(\mathbf{N}_q)_{st}$ is finite for all s and t, so c_{st} is finite too. From the symmetry of $\mathbf{N}_q$ and a standard result concerning cofactors we have

$$(\mathbf{N}_q^{-1})_{1t} = c_{t1} / |\mathbf{N}_q|, \quad t = 1, \cdots, q+1.$$

Then

$$\mathbf{e}_1^T \mathbf{N}_q^{-1} \mathbf{T}_q \mathbf{N}_q^{-1} \mathbf{e}_1 = |\mathbf{N}_q|^{-2} \sum_{s=1}^{q+1} \sum_{t=1}^{q+1} c_{s1} c_{t1} (\mathbf{T}_q)_{st} = O(1),$$

because the leading term in $\mathbf{T}_q$ is $(\mathbf{T}_q)_{11} = O(1)$.

Also note that under A1.6, $f(x_j)^{-1}$ is finite. So

$$\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b} \mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right) = O\left(\frac{1}{n^2h}\right). \tag{1.51}$$

(iii) Thirdly, we will calculate $\frac{1}{n^2}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l} b_b(x_j)b_b(x_l)$ in (1.43). Using the result in (1.49), we get

$$\frac{1}{n^2}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l} b_b(x_j)b_b(x_l) = \frac{1}{n^2}n(n-1)O\left(h^{2q+2}\right) = O\left(h^{2q+2}\right). \tag{1.52}$$

(iv) The last term on the right-hand side of (1.43) is $\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right)$, and

$$\begin{aligned}\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right) &= \frac{\sigma_{NP}^2}{n^2}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}\mathbf{s}_{bj}\boldsymbol{\Omega}_b\mathbf{s}_{bl}^T \\ = \frac{\sigma_{NP}^2}{n^2}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l} &\mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bl}^T\mathbf{X}_{bl}(\mathbf{X}_{bl}^T\mathbf{W}_{bl}\mathbf{X}_{bl})^{-1}\mathbf{e}_1\end{aligned}$$

Note that again,

$$\left|\left(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bl}^T\mathbf{X}_{bl}\right)_{st}\right| \leq \max\{\omega_i\}\left|\left(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{W}_{bl}^T\mathbf{X}_{bl}\right)_{st}\right|,$$

So $\mathbf{X}_{bj}^T\mathbf{W}_{bj}\boldsymbol{\Omega}_b\mathbf{W}_{bl}^T\mathbf{X}_{bl}$ and $\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{W}_{bl}^T\mathbf{X}_{bl}$ have the same order.

Similar to what we have shown before,

$$n^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{W}_{bl}^T\mathbf{X}_{bl} = h^{-1}f(x_j)^{-1}\mathbf{A}\mathbf{T}_{qjl}^*\mathbf{A} + o(h^{-1}\mathbf{A}\mathbf{1}\mathbf{A}), \tag{1.53}$$

where $\mathbf{T}_{qjl}^*$ is the $(q+1)\times(q+1)$ matrix whose (s,t)th entry is

$$H_s * H_t\left(\frac{x_j - x_l}{h}\right),$$

and $H_s * H_t\left(\frac{x_j - x_l}{h}\right)$ is the convolution kernel function of $H_s\left(\frac{x_i - x_j}{h}\right)$ and $H_t\left(\frac{x_i - x_l}{h}\right)$ where

$$H_s\left(\frac{x_i - x_j}{h}\right) \equiv K_{ij}\left(\frac{x_i - x_j}{h}\right)^{s-1}.$$

Now we will show how to obtain (1.53). Note that the (s,t)th entry of $n^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{W}_{bl}^T\mathbf{X}_{bl}$ is

$$\begin{aligned}
&(n^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{W}_{bl}^T\mathbf{X}_{bl})_{st}\\
=\ & \frac{1}{nh^2}\sum_{i\in S_b} K_{ij}K_{ij}(x_i-x_j)^{s-1}(x_i-x_j)^{t-1}\\
=\ & h^{-1}h^{s+t-2}\int K(u)u^{s-1}K\left(u+\frac{x_j-x_l}{h}\right)\left(u+\frac{x_j-x_l}{h}\right)^{t-1} f(x_j)du\\
&+o(h^{-1}h^{s+t-2})\\
=\ & h^{-1}f(x_j)h^{s+t-2}\int H_s(u)H_t\left(u+\frac{x_j-x_l}{h}\right)du+o(h^{-1}h^{s+t-2})\\
=\ & h^{-1}f(x_j)h^{s+t-2}H_s*H_t\left(\frac{x_j-x_l}{h}\right)+o(h^{-1}h^{s+t-2}),
\end{aligned}$$

thus (1.53) holds.

Then using similar reasoning to what was done for (ii), ignoring the terms of small orders, we have

$$\begin{aligned}
&\frac{\sigma_{NP}^2}{n^2}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}\mathbf{e}_1^T(\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{X}_{bj})^{-1}\mathbf{X}_{bj}^T\mathbf{W}_{bj}\mathbf{W}_{bl}^T\mathbf{X}_{bl}(\mathbf{X}_{bl}^T\mathbf{W}_{bl}\mathbf{X}_{bl})^{-1}\mathbf{e}_1\\
\doteq\ & \frac{\sigma_{NP}^2}{n^3}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}\left[f(x_j)^{-1}\mathbf{e}_1^T\mathbf{N}_q^{-1}\mathbf{A}^{-1}h^{-1}f(x_j)\mathbf{A}\mathbf{T}_{qjl}^*\mathbf{A}f(x_l)^{-1}\mathbf{A}^{-1}\mathbf{N}_q^{-1}\mathbf{e}_1\right]\\
=\ & \frac{\sigma_{NP}^2}{n^3}\sum_{j\in S_b}\sum_{l\in S_b, j\neq l} h^{-1}f(x_l)^{-1}\mathbf{e}_1^T\mathbf{N}_q^{-1}\mathbf{T}_{qjl}^*\mathbf{N}_q^{-1}\mathbf{e}_1.
\end{aligned}$$

Similar to what was done in (ii), note that the leading term in $\mathbf{T}_{qjl}^*$ is $(\mathbf{T}_{qjl}^*)_{11}=O(1)$, so

$$\mathbf{e}_1^T\mathbf{N}_q^{-1}\mathbf{T}_{qjl}^*\mathbf{N}_q^{-1}\mathbf{e}_1=|\mathbf{N}_q|^{-2}\sum_{s=1}^{q+1}\sum_{t=1}^{q+1}c_{1s}c_{1t}(\mathbf{T}_{qjl}^*)_{st}=O(1).$$

Therefore,

$$\frac{1}{n^2}\mathrm{E}\left(\sum_{j\in S_b}\sum_{l\in S_b, j\neq l}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right)=O\left(\frac{1}{nh}\right). \tag{1.54}$$

Assumption A1.7 implies that $nh\to\infty$, and by (1.50), (1.51), (1.52) and (1.54),

$$\mathrm{E}(a1)\ =\ O\left(\frac{h^{2q+2}}{n}\right)+O\left(\frac{1}{n^2h}\right)+O\left(h^{2q+2}\right)+O\left(\frac{1}{nh}\right)$$

$$= \quad O\left(h^{2q+2}\right) + O\left(\frac{1}{nh}\right). \tag{1.55}$$

Similarly, we can calculate $\mathrm{E}(a2)$ and $\mathrm{E}(a3)$. Assume A1.7, $nh \to \infty$, so

$$\begin{aligned}
\mathrm{E}(a2) &= \frac{1}{N^2}\sum_{j\in U} b_b^2(x_j) + \frac{1}{N^2}\mathrm{E}\left(\sum_{j\in U}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bj}^T\right) + \frac{1}{N^2}\sum_{j\in U}\sum_{l\in U, j\neq l} b_b(x_j)b_b(x_l) \\
&\quad + \frac{1}{N^2}\mathrm{E}\left(\sum_{j\in U}\sum_{l\in U, j\neq l}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right) \\
&= O\left(\frac{h^{2q+2}}{N}\right) + O\left(\frac{1}{N^2 h}\right) + O\left(h^{2q+2}\right) + O\left(\frac{1}{Nh}\right) \\
&= O\left(h^{2q+2}\right) + O\left(\frac{1}{Nh}\right),
\end{aligned} \tag{1.56}$$

and

$$\begin{aligned}
\mathrm{E}(a3) &= -\frac{2}{nN}\sum_{j\in S_b}\sum_{l\in U} b_b(x_j)b_b(x_l) - \frac{2}{nN}\mathrm{E}\left(\sum_{j\in S_b}\sum_{l\in U}\mathbf{s}_{bj}\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T\mathbf{s}_{bl}^T\right) \\
&= -\frac{2}{nN}\sum_{j\in S_b}\sum_{l\in U} O\left(h^{2q+2}\right) - \frac{2}{nN}\sum_{j\in S_b}\sum_{l\in U} O\left(\frac{1}{nh}\right) \\
&= -\frac{2}{nN}nN\cdot O\left(h^{2q+2}\right) - \frac{2}{nN}nN\cdot O\left(\frac{1}{nh}\right) \\
&= O\left(h^{2q+2}\right) + O\left(\frac{1}{nh}\right).
\end{aligned} \tag{1.57}$$

Also note that $(A) = \frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) \geq 0$, so $|\frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m})| = \frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m})$. Thus, by (1.55), (1.56) and (1.57),

$$\begin{aligned}
\mathrm{E}\left|\frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m})\right| &= \mathrm{E}\left(\frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m})\right) \\
&= \frac{1}{k}\sum_{b=1}^{k}\left\{\mathrm{E}(a1) + \mathrm{E}(a2) + \mathrm{E}(a3)\right\} \\
&= O\left(h^{2q+2}\right) + O\left(\frac{1}{nh}\right).
\end{aligned} \tag{1.58}$$

By Corollary A.2,

$$\frac{1}{kn^2}(\hat{\mathbf{m}}_b - \mathbf{m})^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) = O_p\left(h^{2q+2}\right) + O_p\left(\frac{1}{nh}\right). \tag{1.59}$$

Next,

$$
\begin{aligned}
(B) &= \frac{1}{kn^2}\mathbf{m}^T\mathbf{D}(\hat{\mathbf{m}}_b - \mathbf{m}) \\
&= \frac{1}{k}\sum_{b=1}^{k}\left[\frac{1}{n}\sum_{j\in S_b} m(x_j) - \frac{1}{N}\sum_{j\in U} m(x_j)\right] \\
&\quad \cdot \left[\frac{1}{n}\sum_{j\in S_b}(\hat{m}(x_j) - m(x_j)) - \frac{1}{N}\sum_{j\in U}(\hat{m}(x_j) - m(x_j))\right] \\
&\equiv \frac{1}{k}\sum_{b=1}^{k}\Delta_{1S_b}\check{\Delta}_{2S_b}.
\end{aligned}
$$

Since $m(x_j)$'s are fixed numbers, and $\Delta_{1S_b} = \frac{1}{n}\sum_{j\in S_b} m(x_j) - \frac{1}{N}\sum_{j\in U} m(x_j) = O(1)$,

$$
\begin{aligned}
\mathrm{E}(B)^2 &= \mathrm{E}\left(\frac{1}{k^2}\sum_{b=1}^{k}\sum_{c=1}^{k}\Delta_{1S_b}\check{\Delta}_{2S_b}\Delta_{1S_c}\check{\Delta}_{2S_c}\right) \\
&= \frac{1}{k^2}\sum_{b=1}^{k}\sum_{c=1}^{k}\Delta_{1S_b}\Delta_{1S_c}\mathrm{E}\left(\check{\Delta}_{2S_b}\check{\Delta}_{2S_c}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{E}\left(\check{\Delta}_{2S_b}\check{\Delta}_{2S_c}\right) &= \mathrm{E}\left\{\left[\frac{1}{n}\sum_{j\in S_b}(\hat{m}(x_j) - m(x_j)) - \frac{1}{N}\sum_{j\in U}(\hat{m}(x_j) - m(x_j))\right]\right. \\
&\quad \left.\cdot \left[\frac{1}{n}\sum_{j\in S_c}(\hat{m}(x_j) - m(x_j)) - \frac{1}{N}\sum_{j\in U}(\hat{m}(x_j) - m(x_j))\right]\right\} \\
&= \mathrm{E}\left\{\frac{1}{n}\sum_{j\in S_b}(\hat{m}(x_j) - m(x_j))\,\frac{1}{n}\sum_{j\in S_c}(\hat{m}(x_j) - m(x_j))\right. \\
&\quad -\frac{1}{nN}\sum_{j\in S_b}(\hat{m}(x_j) - m(x_j))\sum_{j\in U}(\hat{m}(x_j) - m(x_j)) \\
&\quad -\frac{1}{nN}\sum_{j\in S_c}(\hat{m}(x_j) - m(x_j))\sum_{j\in U}(\hat{m}(x_j) - m(x_j)) \\
&\quad \left.+\frac{1}{N^2}\left[\sum_{j\in U}(\hat{m}(x_j) - m(x_j))\right]^2\right\}.
\end{aligned}
$$

The reasoning for $\mathrm{E}\left(\check{\Delta}_{2S_b}\check{\Delta}_{2S_c}\right)$ is analogous to what was done for $\mathrm{E}\left[(a1) + (a2) + (a3)\right]$, so

$$
\mathrm{E}\left(\check{\Delta}_{2S_b}\check{\Delta}_{2S_c}\right) = O\left(h^{2q+2}\right) + O\left(\frac{1}{nh}\right).
$$

Thus,

$$\begin{aligned}
\mathrm{E}(B)^2 &= \frac{1}{k^2}\sum_{b=1}^{k}\sum_{c=1}^{k}\Delta_{1S_b}\Delta_{1S_c}\mathrm{E}\left(\check{\Delta}_{2S_b}\check{\Delta}_{2S_c}\right)\\
&= \frac{1}{k^2}\sum_{b=1}^{k}\sum_{c=1}^{k}\Delta_{1S_b}\Delta_{1S_c}\left(O\left(h^{2q+2}\right)+O\left(\frac{1}{nh}\right)\right)\\
&= O\left(h^{2q+2}\right)+O\left(\frac{1}{nh}\right). \qquad (1.60)
\end{aligned}$$

By Corollary A.1,

$$(B) = O_p\left(h^{q+1}\right)+O_p\left(\frac{1}{\sqrt{nh}}\right). \qquad (1.61)$$

Therefore,

$$\begin{aligned}
\frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{D}\hat{\mathbf{m}}_b-\mathbf{m}^T\mathbf{D}\mathbf{m}) &= (A)+2(B)\\
&= O_p\left(h^{2q+2}\right)+O_p\left(\frac{1}{nh}\right)+O_p\left(h^{q+1}\right)+O_p\left(\frac{1}{\sqrt{nh}}\right)\\
&= O_p\left(h^{q+1}\right)+O_p\left(\frac{1}{\sqrt{nh}}\right). \qquad (1.62)
\end{aligned}$$

Now Let us evaluate the second term on the right hand side of (1.40), which is $\frac{1}{kn^2}\mathrm{tr}(\mathbf{D}\mathbf{\Omega})(\hat{\sigma}^2_{NPb}-\sigma^2_{NP})$. Note that

$$\begin{aligned}
\hat{\sigma}^2_{NPb}-\sigma^2_{NP} &= \frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-\hat{m}(x_j))^2}{\omega_j}-\sigma^2_{NP}\\
&= \frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j)+m(x_j)-\hat{m}(x_j))^2}{\omega_j}-\sigma^2_{NP}\\
&= \frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))^2}{\omega_j}-\sigma^2_{NP}+\frac{1}{n}\sum_{j\in S_b}\frac{(\hat{m}(x_j)-m(x_j))^2}{\omega_j}\\
&\quad +\frac{2}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))(\hat{m}(x_j)-m(x_j))}{\omega_j}. \qquad (1.63)
\end{aligned}$$

Let us evaluate the first two terms on the right hand side of (1.63). Note that

$$\mathrm{E}\left(\frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))^2}{\omega_j}-\sigma^2_{NP}\right) = \mathrm{E}\left(\frac{1}{n}\sum_{j\in S_b}\frac{\varepsilon_j^2}{\omega_j}-\sigma^2_{NP}\right)=0$$

and

$$\operatorname{Var}\left(\frac{1}{n}\sum_{j\in S_b}\frac{\varepsilon_j^2}{\omega_j}-\sigma_{NP}^2\right) = \frac{\operatorname{Var}(\varepsilon_j^2/\omega_j)}{n},$$

where $\operatorname{Var}(\varepsilon_j^2/\omega_j)$ is bounded. By Corollary A.1,

$$\frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))^2}{\omega_j}-\sigma_{NP}^2 = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{1.64}$$

The third term on the right side of (1.63) is $\frac{1}{n}\sum_{j\in S_b}(\hat{m}(x_j)-m(x_j))^2/\omega_j$ and it is the same as the first term on the right-hand side of (1.41) except for a factor of n. The bounded number ω_j in the denominator inside the summation does not affect the order in probability. So using similar techniques,

$$\frac{1}{n}\sum_{j\in S_b}\frac{(\hat{m}(x_j)-m(x_j))^2}{\omega_j} = O_p\left(h^{2q+2}\right)+O_p\left(\frac{1}{nh}\right). \tag{1.65}$$

For the last term on the right side of (1.63), note that

$$\begin{aligned}
&\frac{2}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))(\hat{m}(x_j)-m(x_j))}{\omega_j}\\
\leq\ & 2\left|\frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))(\hat{m}(x_j)-m(x_j))}{\omega_j}\right|\\
\leq\ & 2\sqrt{\frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))^2}{\omega_j}\frac{1}{n}\sum_{j\in S_b}\frac{(\hat{m}(x_j)-m(x_j))^2}{\omega_j}},
\end{aligned}$$

where

$$\frac{1}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))^2}{\omega_j} = \frac{1}{n}\sum_{j\in S_b}\frac{\varepsilon^2}{\omega_j} = O_p(1)$$

and

$$\frac{1}{n}\sum_{j\in S_b}\frac{(\hat{m}(x_j)-m(x_j))^2}{\omega_j} = O_p\left(h^{2q+2}\right)+O_p\left(\frac{1}{nh}\right),$$

so

$$\frac{2}{n}\sum_{j\in S_b}\frac{(Y_j-m(x_j))(\hat{m}(x_j)-m(x_j))}{\omega_j} = O_p\left(h^{q+1}\right)+O_p\left(\frac{1}{\sqrt{nh}}\right). \tag{1.66}$$

Since $\frac{1}{kn^2}\text{tr}(\mathbf{D\Omega}) = O(1)$, and by (1.64), (1.65) and (1.66),

$$\begin{aligned} & \frac{1}{kn^2}\text{tr}(\mathbf{D\Omega})(\hat{\sigma}^2_{NPb} - \sigma^2_{NP}) \\ = \;& O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(h^{2q+2}\right) + O_p\left(\frac{1}{nh}\right) + O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right) \\ = \;& O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right). \end{aligned} \tag{1.67}$$

Therefore by (1.62) and (1.67),

$$\begin{aligned} \hat{\text{V}}_{NP}(\bar{Y}_S) - \text{E}_{NP}[\text{Var}_p(\bar{Y}_S)] &= O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right) + O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= O_p\left(h^{q+1}\right) + O_p\left(\frac{1}{\sqrt{nh}}\right), \end{aligned} \tag{1.68}$$

and equation (1.37) holds.

$\square$

1.5 Simulation Study

To further investigate the statistical properties of the above variance estimators, we perform a simulation study. For simplicity, we consider the case where there is only one auxiliary variable x. We also assume that the errors are independently and normally distributed with homogeneous variances. Two superpopulation models are examined. One is the linear model

$$y_j = 5 + 2x_j + \varepsilon_j, \tag{1.69}$$

where $j = 1, \ldots, N$ and $\varepsilon_j \sim NID(0, \sigma_1^2)$. The other is the quadratic model

$$y_j = 5 + 2x_j - 2x_j^2 + \varepsilon_j, \tag{1.70}$$

where $j = 1, \ldots, N$ and $\varepsilon_j \sim NID(0, \sigma_2^2)$.

Let R_1^2 and R_2^2 denote the *coefficient of determination* for model (1.69) and (1.70), respectively. The coefficient of determination, also known as R-square, is the fraction of variation in the response that is explained by the model. It is defined as

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i \in U} \varepsilon_i^2}{\sum_{i \in U} (Y_i - \bar{Y}_U)^2}.$$

Bigger R-square means bigger predictive power of the model. We vary the value of σ_1 and σ_2 to achieve high and low R_1^2 and R_2^2, respectively. Specifically, we let $R^2 \approx 0.75$ and $R^2 \approx 0.25$ for each of (1.69) and (1.70).

To draw a systematic sample, we first need to sort the population. We consider three ways: (1) Sort by auxiliary variable x; (2) Sort by z_1, where $z_{1j} = x_j + \eta_{1j}$ and $\eta_{1j} \sim NID(0, \sigma_{\eta_1}^2)$. Choose $\sigma_{\eta_1}^2$ to make $R_{z_1}^2 = 0.75$. (3) Sort by z_2, where $z_{2j} = x_j + \eta_{2j}$ and $\eta_{2j} \sim NID(0, \sigma_{\eta_2}^2)$. Choose $\sigma_{\eta_2}^2$ to make $R_{z_2}^2 = 0.25$.

We generate populations of size $N = 2000$. To achieve this, we generate 2000 values of model variable x from the uniform distribution on $[0, 1]$ and 2000 values of error ε from $N(0, 1)$, up to multiplication by σ_1 or σ_2. Then we compute 2000 values of response variable y by model (1.69) and (1.70). We consider two systematic samples of size $n = 500$ and $n = 100$, with corresponding sampling intervals $k = 4$ and $k = 20$, respectively. To draw a systematic sample, we first sort the data, either by x or z, from the smallest to the largest, then we randomly choose an observation from the first k observations, say the bth one. Then our sample consists of the observations with the following subscripts: b, $b + k$, ..., $b + (n - 1)k$. For each simulation, we calculate the corresponding $\mathrm{Var}_p(\bar{Y}_S)$, $\mathrm{E}_M[\mathrm{Var}_p(\bar{Y}_S)]$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, as defined in (1.6), (1.34), (1.35), (1.9), (1.3), (1.4) and (1.2), respectively. For $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, we use local linear regression to compute $\hat{m}(x_j)$, with the following kernel function:

$$K(t) = \begin{cases} 1 - t^2, & |t| \leq 1; \\ 0 & \text{Otherwise.} \end{cases} \tag{1.71}$$

We consider three bandwidth values: $h = 0.50$, $h = 0.25$ and $h = 0.10$. Each simulation setting is repeated $B = 10000$ times.

In summary, we study 24 scenarios (four superpopulation models, three sorting criteria and two sample sizes) for $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ and 72 scenarios (three bandwidth values in addition) for $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$.

We compare the performance of the nonparametric variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ with $\hat{\mathrm{V}}_L(\bar{Y}_S)$, the linear variance estimator proposed by Bartolucci and Montanari (2006), and with the overlapping differences estimator $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and the nonoverlapping differences estimator $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$, which are recommended by Wolter (1985). The expressions of $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ are defined in (1.3) and (1.4), respectively. We also compare $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ with $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, the variance estimator for simple random sampling (SRS) because $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ has the simplest form. It is defined as in (1.2).

Let us use $\hat{\mathrm{V}}(\bar{Y}_S)$ as a generic notation for $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$. We calculate the relative bias (RB), mean squared error (MSE) and mean squared prediction error (MSPE), where

$$\begin{aligned} \mathrm{RB} &= \frac{\mathrm{E}_{Mp}(\hat{\mathrm{V}}(\bar{Y}_S)) - \mathrm{E}_M[\mathrm{Var}_p(\bar{Y}_S)]}{\mathrm{E}_M[\mathrm{Var}_p(\bar{Y}_S)]}, \\ \mathrm{MSE} &= \mathrm{E}_{Mp}(\hat{\mathrm{V}}(\bar{Y}_S) - \mathrm{E}_M[\mathrm{Var}_p(\bar{Y}_S)])^2, \\ \text{and} \quad \mathrm{MSPE} &= \mathrm{E}_{Mp}(\hat{\mathrm{V}}(\bar{Y}_S) - \mathrm{Var}_p(\bar{Y}_S))^2. \end{aligned}$$

In the above expressions, E_M denotes the expectation under the superpopulation model M, and E_{Mp} denotes the expectation under both the model and design. Note that in this model-based context, $\mathrm{Var}_p(\bar{Y}_S)$ is a random variable, so $\hat{\mathrm{V}}(\bar{Y}_S)$ is a predictor rather than an estimator for it.

It is useful to investigate the Mean Squared Error (MSE) of the above variance estimators because MSE measures the variability of each variance estimator. Similarly, Mean Squared Prediction Error (MSPE) is a useful measurement of the variability of each variance estimator as a predictor for $\mathrm{Var}_p(\bar{Y}_S)$.

Table 1.2 reports the relative biases (in percent) of $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ (evaluated at three bandwidth values), $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, when populations are sorted by auxiliary variable x. Smaller relative bias values are more desirable. We can see that $\hat{\mathrm{V}}_L(\bar{Y}_S)$ performs very well when the superpopulation model is linear. It is almost unbiased. However, when the superpopulation model is not linear, its bias increases dramatically due to the misspecification. The nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ performs well under all four superpopulation models if a proper bandwidth value is chosen. Specifically, when the superpopulation model is linear, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ tends to favor bigger bandwidth. This is because we used local linear regression in the calculation of $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$. Specifically, local polynomial regression with kernel function (1.71) uses the points that are within an interval of length $2h$ around point x_j. Bigger bandwidth results in more points in that interval. And because our local polynomial regression is local linear, which is the correct one for this population with linear trend, so having more points will increase the precision of each local linear regression. When the superpopulation is quadratic, it tends to favor smaller bandwidth. The reasons for this fact are as follows. As we have shown, for parametric estimation, linear regression will not estimate the quadratic trend well. The wider the interval, the more likely we will see a quadratic trend there. Therefore, local linear regression on that interval could be bad. For example, the bias for $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ with $h = 0.5$ and $R_2^2 \approx 0.75$ is more than 56%. If the bandwidth is smaller, then the trend within each local interval will be better approximated by a linear trend. Note that there is a tradeoff between using smaller bandwidth and using a correct regression within each local interval. In this paper we will not further discuss the bandwidth selection problem. We choose these three bandwidth values to illustrate the bandwidth effect. See Opsomer and Miller (2005), for example, for more information on optimal bandwidth selection. We also see that $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ have small biases under all four models. This is because when the populations are sorted by x before drawing systematic samples, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ can capture the population trend very well and thus very efficient. The most inefficient variance estimator in this case is $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$. It always overestimates the true variance.

Table 1.3 reports the relative biases (in percent) of $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ (evaluated at three bandwidth values), $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, when populations are sorted by variable z_1, where z_1 is highly correlated with x. For $\hat{\mathrm{V}}_L(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, we can draw similar conclusions to those in Table 1.2. The linear estimator $\hat{\mathrm{V}}_L(\bar{Y}_S)$ does well when the superpopulation model is linear, but has large biases when the superpopulation model is quadratic. However, under the quadratic superpopulation model, its relative biases are smaller than those in Table 1.2. The nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ again performs well if we choose a proper bandwidth. The most important difference we can see is that $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ have larger bias values than those in Table 1.2. This is because when we sort the populations by z_1, the overlapping and nonoverlapping difference estimator $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ cannot capture the population trend as well as the previous case. But one thing we need to note is that as shown in (1.3) and (1.4),

$\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ do not depend on auxiliary variable x, so our nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ has the advantage of using auxiliary information to improve its precision under this circumstance.

Table 1.4 reports the relative biases (in percent) of $\hat{\mathrm{V}}_{L}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ (evaluated at three bandwidth values), $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ when populations are sorted by variable z_2, where z_2 has low correlation with x. We can see a similar trend as in Table 1.2 and Table 1.3. That is, when the superpopulation model is linear, the linear estimator $\hat{\mathrm{V}}_{L}(\bar{Y}_S)$ performs well. When the superpopulation model deviates from linearity, it becomes more biased. Given a proper bandwidth, our nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ has small bias for all four superpopulation models and both sample sizes. We can also see that the overlapping difference estimator $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and nonoverlapping difference estimator $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ tend to have more biases than those in Table 1.2. Note that this is the case where the populations are sorted by z_2, which result in almost random permutations of populations. So the systematic samples are close to SRS samples. Thus it is not surprising to see that $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ have similar bias with $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, especially when the superpopulation model is quadratic.

Table 1.5 reports the MSEs of $\hat{\mathrm{V}}_{L}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ (evaluated at three bandwidth values), $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ when populations are sorted by model variable x. To simplify the illustration, we use $\hat{\mathrm{V}}_{NP}(\bar{Y}_S), h = 0.10$ as the target for comparisons. We list the ratios of MSEs for all other variance estimators and $\mathrm{MSE}(\hat{\mathrm{V}}_{NP}(\bar{Y}_S), h = 0.10)$. If a ratio is less than one, it means that variance estimator on the numerator has less variability than $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ with $h = 0.10$. We can see that when the superpopulation model is linear, the linear estimator $\hat{\mathrm{V}}_{L}(\bar{Y}_S)$ and our nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ with larger bandwidth values tend to be favorable choices. When the superpopulation model is quadratic, our nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ performs the best if given a proper bandwidth. This can be seen by noting that in the last two columns of Table 1.5, only three values are less than one, and they are the ratios of our nonparametric estimator at different bandwidth values. In all cases, although never the best, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ do not perform too badly either. They are close to the best choice in each case. The linear estimator $\hat{\mathrm{V}}_{L}(\bar{Y}_S)$ drastically fails when the superpopulation model is quadratic. The SRS variance estimator $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ is almost always a bad choice.

Table 1.6 and Table 1.7 report the same ratios as in Table 1.5, except that in these two tables, populations are sorted by z_1 and z_2, respectively. We see a similar trend as in Table 1.5. What is different here is that as the sorting variable z deviates further from the model variable x, the overlapping difference estimator $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and nonoverlapping difference estimator $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ tend to be worse. Note that in Table 1.6, when the superpopulation model is quadratic and the sample size is 500, the linear estimator and SRS estimator are not too bad at all. This is because we generate sorting variables z_1 and z_2 only once, and use the same sequence of z_1 and z_2 in each replication. This is to achieve the same permutation of model function $m(x)$ so that we will have the same $\mathrm{E}_M[\mathrm{Var}_p(\bar{Y}_S)]$ for all 10000 replications. Thus the unusually low values may happen by chance and be true for only this particular set of z_1 and z_2.

Table 1.8, Table 1.9 and Table 1.10 report the MSPEs of $\hat{\mathrm{V}}_{L}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ at $h =$ 0.50 and 0.25, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, relative to the MSPE for $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ at $h = 0.10$, where populations are sorted by x, z_1 and z_2, respectively. We can see that our nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is again the overall best choice if given a proper bandwidth. However, the differences between different variance estimators in terms of MSPEs are much smaller than those of MSE. This fact indicates that all variance estimators have similar variability in terms of predicting the design variance $\mathrm{Var}_p(\bar{Y}_S)$.

1.6 Simulation conclusions

Summarizing the above statements, we came to the following conclusions about our simulation study.

1. The linear estimator $\hat{\mathrm{V}}_L(\bar{Y}_S)$ has low bias, small MSE and small MSPE when the superpopulation model is linear. However, when the superpopulation model departs from linearity, it can drastically fail.

2. Our nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is overall the best choice if given a proper bandwidth. In terms of relative bias, it is always one of the best choices. Other variance estimators may have smaller bias than $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ under certain circumstances, for example, the linear estimator $\hat{\mathrm{V}}_L(\bar{Y}_S)$ under linear superpopulation models and the overlapping difference estimator $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and nonoverlapping difference estimator $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ when the populations are sorted by model variable x. However, those estimators are not consistently good for all cases. Our nonparametric estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is robust to superpopulation models and sorting variables. As far as MSE is concerned, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ always outperforms $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$. It may occasionally be worse than $\hat{\mathrm{V}}_L(\bar{Y}_S)$ under linear superpopulation models, but as we have mentioned before, $\hat{\mathrm{V}}_L(\bar{Y}_S)$ is usually bad for nonlinear superpopulation models.

3. The overlapping difference estimator $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$ and nonoverlapping difference estimator $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ perform similarly. They both have very small bias when the populations are sorted by x. However, as the sorting variable deviates further from model variable x, they tend to have bigger bias. In terms of MSE and MSPE, they are never the best choices.

4. The SRS estimator $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ can be good for certain cases, such as when the populations are sorted by z_2. However, it is overall not recommended because in most cases it behaves poorly.

5. The differences among different estimators with respect to MSPE are much smaller than those of MSE.

6. If one is to use the nonparametric variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ in a real survey problem, we recommend using smaller bandwidth values rather than bigger ones. As we have shown, although $h = 0.10$ may not be the best choices in terms of MSE, it is usually the best choice with respect to relative bias.

Relative Bias (%)		Linear: 1: $R_1^2 \approx 0.75$	Linear: 2: $R_1^2 \approx 0.25$	Quadratic: 3: $R_2^2 \approx 0.75$	Quadratic: 4: $R_2^2 \approx 0.25$
$\hat{\mathrm{V}}_L(\bar{Y}_S)$	$n = 500$	-0.04	-0.08	329.00	32.85
	$n = 100$	0.02	0.28	334.61	33.51
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.50$	$n = 500$	-0.66	-0.69	58.48	5.31
	$n = 100$	-2.95	-3.38	56.33	2.85
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.25$	$n = 500$	-1.01	-1.05	5.32	-0.33
	$n = 100$	-4.63	-5.12	1.85	-4.19
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.10$	$n = 500$	-2.05	-2.10	-1.88	-2.00
	$n = 100$	-9.66	-10.25	-9.49	-10.00
$\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$	$n = 500$	-0.86	-0.13	-0.06	-0.03
	$n = 100$	-3.04	-0.13	0.98	-0.04
$\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$	$n = 500$	-0.83	-0.19	-0.04	-0.06
	$n = 100$	-3.16	-0.06	0.92	-0.18
$\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$	$n = 500$	330.93	33.24	328.41	32.80
	$n = 100$	322.05	33.73	331.42	33.18

Table 1.2: Simulated relative bias (in percent) for $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by x.

Relative Bias (%)		Linear 1: $R_1^2 \approx 0.75$	Linear 2: $R_1^2 \approx 0.25$	Quadratic 3: $R_2^2 \approx 0.75$	Quadratic 4: $R_2^2 \approx 0.25$
$\hat{\mathrm{V}}_L(\bar{Y}_S)$	$n = 500$	0.01	0.09	7.95	2.38
	$n = 100$	0.26	0.22	72.78	16.20
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.50$	$n = 500$	-0.09	-0.43	-23.66	-7.61
	$n = 100$	0.25	-1.98	-6.81	-3.55
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.25$	$n = 500$	-0.19	-0.60	-9.42	-3.37
	$n = 100$	-0.07	-2.69	-2.59	-3.20
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.10$	$n = 500$	-0.47	-1.03	-1.60	-1.30
	$n = 100$	0.36	-1.89	2.41	-1.54
$\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$	$n = 500$	10.92	1.74	-23.94	-7.42
	$n = 100$	-18.22	-3.63	20.38	4.68
$\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$	$n = 500$	10.67	1.78	-24.49	-7.62
	$n = 100$	-17.37	-3.35	21.77	5.01
$\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$	$n = 500$	171.18	25.93	7.80	2.31
	$n = 100$	98.30	19.19	71.97	15.74

Table 1.3: Simulated relative bias (in percent) for $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by z_1.

Relative Bias (%)		Linear 1: $R_1^2 \approx 0.75$	Linear 2: $R_1^2 \approx 0.25$	Quadratic 3: $R_2^2 \approx 0.75$	Quadratic 4: $R_2^2 \approx 0.25$
$\hat{\mathrm{V}}_L(\bar{Y}_S)$	$n = 500$	0.04	0.10	163.95	25.10
	$n = 100$	0.39	0.49	-4.37	-1.04
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.50$	$n = 500$	0.08	-0.31	17.37	2.30
	$n = 100$	0.85	-1.28	-26.06	-9.88
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.25$	$n = 500$	-0.01	-0.33	-3.41	-0.90
	$n = 100$	0.80	-1.42	-6.05	-3.35
$\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $h = 0.10$	$n = 500$	-0.07	-0.29	0.21	-0.36
	$n = 100$	1.25	-0.13	3.07	0.85
$\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$	$n = 500$	15.44	3.85	156.26	23.76
	$n = 100$	9.32	2.47	-4.62	-1.58
$\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$	$n = 500$	13.93	3.51	158.33	24.05
	$n = 100$	13.72	3.72	-5.41	-1.72
$\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$	$n = 500$	45.73	11.36	163.90	24.99
	$n = 100$	33.80	8.78	-4.90	-1.68

Table 1.4: Simulated relative bias (in percent) for $\hat{\mathrm{V}}_L(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{OL}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by z_2.

		Linear		Quadratic	
		1: $R_1^2 \approx 0.75$	2: $R_1^2 \approx 0.25$	3: $R_2^2 \approx 0.75$	4: $R_2^2 \approx 0.25$
$\frac{\text{MSE}(\hat{\text{V}}_L(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.91	0.92	2544.64	26.09
	$n = 100$	0.74	0.73	423.22	5.21
$\frac{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.50)}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.93	0.92	81.32	1.59
	$n = 100$	0.72	0.71	12.70	0.73
$\frac{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.25)}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.94	0.93	1.59	0.92
	$n = 100$	0.76	0.75	0.72	0.74
$\frac{\text{MSE}(\hat{\text{V}}_{OL}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.39	1.36	1.40	1.37
	$n = 100$	1.14	1.10	1.13	1.08
$\frac{\text{MSE}(\hat{\text{V}}_{NO}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.84	1.82	1.88	1.81
	$n = 100$	1.49	1.43	1.53	1.45
$\frac{\text{MSE}(\hat{\text{V}}_{SRS}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	2560.16	27.02	2535.50	26.00
	$n = 100$	412.34	5.34	415.18	5.11

Table 1.5: Comparisons between MSE of $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.10$ and MSEs of $\hat{\text{V}}_L(\bar{Y}_S)$, $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.50$ and 0.25, $\hat{\text{V}}_{OL}(\bar{Y}_S)$, $\hat{\text{V}}_{NO}(\bar{Y}_S)$ and $\hat{\text{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by x before drawing systematic samples.

		Linear		Quadratic	
		1: $R_1^2 \approx 0.75$	2: $R_1^2 \approx 0.25$	3: $R_2^2 \approx 0.75$	4: $R_2^2 \approx 0.25$
$\dfrac{\text{MSE}(\hat{\text{V}}_L(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.10)}$	$n = 500$	0.40	0.87	2.77	0.86
	$n = 100$	0.70	0.92	41.52	2.65
$\dfrac{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.50)}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.10)}$	$n = 500$	0.54	0.86	16.20	1.64
	$n = 100$	0.73	0.88	0.95	0.88
$\dfrac{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.25)}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.10)}$	$n = 500$	0.70	0.91	3.03	0.92
	$n = 100$	0.80	0.91	0.76	0.90
$\dfrac{\text{MSE}(\hat{\text{V}}_{OL}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.10)}$	$n = 500$	4.15	1.57	17.59	1.98
	$n = 100$	4.74	1.51	6.00	1.77
$\dfrac{\text{MSE}(\hat{\text{V}}_{NO}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.10)}$	$n = 500$	4.56	2.06	18.10	2.33
	$n = 100$	5.11	1.98	6.82	2.30
$\dfrac{\text{MSE}(\hat{\text{V}}_{SRS}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h = 0.10)}$	$n = 500$	615.73	17.18	2.70	0.85
	$n = 100$	89.43	3.28	40.61	2.56

Table 1.6: Comparisons between MSE of $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.10$ and MSEs of $\hat{\text{V}}_L(\bar{Y}_S)$, $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.50$ and 0.25, $\hat{\text{V}}_{OL}(\bar{Y}_S)$, $\hat{\text{V}}_{NO}(\bar{Y}_S)$ and $\hat{\text{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by z_1 before drawing systematic samples.

		Linear		Quadratic	
		1: $R_1^2 \approx 0.75$	2: $R_1^2 \approx 0.25$	3: $R_2^2 \approx 0.75$	4: $R_2^2 \approx 0.25$
$\frac{\text{MSE}(\hat{\text{V}}_L(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.18	0.45	449.56	14.98
	$n = 100$	0.65	0.82	1.11	0.78
$\frac{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.50)}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.25	0.49	5.46	0.92
	$n = 100$	0.72	0.81	5.28	0.96
$\frac{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.25)}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.46	0.63	0.62	0.83
	$n = 100$	0.82	0.87	0.93	0.81
$\frac{\text{MSE}(\hat{\text{V}}_{OL}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	3.55	1.02	410.22	14.16
	$n = 100$	4.14	1.52	2.14	1.18
$\frac{\text{MSE}(\hat{\text{V}}_{NO}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	3.33	1.26	426.71	15.27
	$n = 100$	6.83	2.13	2.54	1.52
$\frac{\text{MSE}(\hat{\text{V}}_{SRS}(\bar{Y}_S))}{\text{MSE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	25.32	2.27	449.14	14.85
	$n = 100$	11.57	1.42	1.13	0.77

Table 1.7: Comparisons between MSE of $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.10$ and MSEs of $\hat{\text{V}}_L(\bar{Y}_S)$, $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.50$ and 0.25, $\hat{\text{V}}_{OL}(\bar{Y}_S)$, $\hat{\text{V}}_{NO}(\bar{Y}_S)$ and $\hat{\text{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by z_2 before drawing systematic samples.

		Linear		Quadratic	
		1: $R_1^2 \approx 0.75$	2: $R_1^2 \approx 0.25$	3: $R_2^2 \approx 0.75$	4: $R_2^2 \approx 0.25$
$\dfrac{\text{MSPE}(\hat{\text{V}}_L(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.98	1.01	18.35	1.18
	$n = 100$	0.92	0.96	86.52	1.93
$\dfrac{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.50)}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.00	1.00	1.57	1.01
	$n = 100$	0.94	0.94	3.38	0.95
$\dfrac{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.25)}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.00	1.00	1.01	1.00
	$n = 100$	0.95	0.95	0.95	0.95
$\dfrac{\text{MSPE}(\hat{\text{V}}_{OL}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.00	1.00	1.00	1.00
	$n = 100$	1.02	1.04	1.02	1.02
$\dfrac{\text{MSPE}(\hat{\text{V}}_{NO}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.01	1.00	1.01	1.01
	$n = 100$	1.09	1.10	1.10	1.10
$\dfrac{\text{MSPE}(\hat{\text{V}}_{SRS}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	17.07	1.17	18.28	1.18
	$n = 100$	77.28	1.91	84.89	1.91

Table 1.8: Comparisons between MSPE of $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.10$ and MSPEs of $\hat{\text{V}}_L(\bar{Y}_S)$, $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.50$ and 0.25, $\hat{\text{V}}_{OL}(\bar{Y}_S)$, $\hat{\text{V}}_{NO}(\bar{Y}_S)$ and $\hat{\text{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by x before drawing systematic samples.

		Linear		Quadratic	
		1: $R_1^2 \approx 0.75$	2: $R_1^2 \approx 0.25$	3: $R_2^2 \approx 0.75$	4: $R_2^2 \approx 0.25$
$\frac{\text{MSPE}(\hat{\text{V}}_L(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.00	1.00	1.02	1.00
	$n = 100$	0.95	0.99	7.84	1.27
$\frac{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.50)}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.00	1.00	1.18	1.00
	$n = 100$	0.96	0.98	0.99	0.98
$\frac{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.25)}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.00	1.00	1.02	1.00
	$n = 100$	0.97	0.98	0.96	0.98
$\frac{\text{MSPE}(\hat{\text{V}}_{OL}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.03	1.00	1.20	1.01
	$n = 100$	1.48	1.08	1.83	1.14
$\frac{\text{MSPE}(\hat{\text{V}}_{NO}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.04	1.01	1.20	1.01
	$n = 100$	1.52	1.16	1.98	1.23
$\frac{\text{MSPE}(\hat{\text{V}}_{SRS}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	6.27	1.10	1.02	1.00
	$n = 100$	12.48	1.36	7.68	1.25

Table 1.9: Comparisons between MSPE of $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.10$ and MSPEs of $\hat{\text{V}}_L(\bar{Y}_S)$, $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.50$ and 0.25, $\hat{\text{V}}_{OL}(\bar{Y}_S)$, $\hat{\text{V}}_{NO}(\bar{Y}_S)$ and $\hat{\text{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by z_1 before drawing systematic samples.

		Linear		Quadratic	
		1: $R_1^2 \approx 0.75$	2: $R_1^2 \approx 0.25$	3: $R_2^2 \approx 0.75$	4: $R_2^2 \approx 0.25$
$\frac{\text{MSPE}(\hat{\text{V}}_L(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.98	0.99	5.54	1.09
	$n = 100$	0.94	0.97	1.03	0.95
$\frac{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.50)}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.98	0.99	1.05	1.00
	$n = 100$	0.96	0.97	2.07	0.99
$\frac{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.25)}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	0.99	1.00	1.00	1.00
	$n = 100$	0.97	0.98	0.98	0.96
$\frac{\text{MSPE}(\hat{\text{V}}_{OL}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.05	1.00	5.14	1.08
	$n = 100$	1.55	1.09	1.29	1.03
$\frac{\text{MSPE}(\hat{\text{V}}_{NO}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.05	1.00	5.31	1.09
	$n = 100$	2.03	1.19	1.40	1.09
$\frac{\text{MSPE}(\hat{\text{V}}_{SRS}(\bar{Y}_S))}{\text{MSPE}(\hat{\text{V}}_{NP}(\bar{Y}_S), h=0.10)}$	$n = 500$	1.53	1.01	5.53	1.09
	$n = 100$	2.87	1.07	1.03	0.95

Table 1.10: Comparisons between MSPE of $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.10$ and MSPEs of $\hat{\text{V}}_L(\bar{Y}_S)$, $\hat{\text{V}}_{NP}(\bar{Y}_S)$ at $h = 0.50$ and 0.25, $\hat{\text{V}}_{OL}(\bar{Y}_S)$, $\hat{\text{V}}_{NO}(\bar{Y}_S)$ and $\hat{\text{V}}_{SRS}(\bar{Y}_S)$, where populations are sorted by z_2 before drawing systematic samples.

CHAPTER 2 Applications of model-based nonparametric variance estimator in Forest Inventory and Analysis (FIA)

2.1 Introduction

In the previous chapter, we have shown the good properties of $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ through theoretical results and a simulation study. In this chapter, we will discuss an application of $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ using real data from *Forest Inventory and Analysis* (FIA).

Forest Inventory and Analysis (FIA) is a program within the U.S. Department of Agriculture Forest Service that conducts nationwide forest surveys (U. S. Department of Agriculture Forest Service, 1992; Frayer and Furnival, 1999; Gillespie, 1999). In these surveys, the population quantities of interest are, for example, total tree volume, growth and mortality, or area by forest type. Design-based estimates of such quantities are produced on a regular basis. The data we are considering are within a 2.5 million ha ecological province (Bailey et al. 1994) that includes the Wasatch and Uinta Mountain Ranges of northern Utah. In the lower elevations, forests in the area generally consist of pinyon-juniper, oak, and maple. In the higher elevations, forest types are generally lodgepole pine, ponderosa pine, aspen, and spruce-fir. Many forest types blend and mix across elevation zones due to other topographic variables such as aspect and slope. In addition to its ecological diversity, the area also has many large ownerships including National Forests, Indian Reservations, National Parks and Monuments, state land holdings, and private land. Each ownership faces different land management issues that require accurate information of forest resource. Figure 2.1 displays the study region and sample points that are collected in early 1990's for the survey we will consider here. Although this paper will focus on this particular example, the general approach can be applied to other regions.

The forest survey data are collected using a two-phase systematic sampling design. In phase one, remote sensing data and geographical information system (GIS) coverage information are gathered on an intensive sample grid. In phase two, a field-visited subset of the phase one grid is taken. Several hundred variables are collected during these field visits, ranging from individual tree characteristics and size measurements to complex ecological health ratings. We will treat the phase one plots roughly as the population of interest and phase two plots as the systematic sample. At the population level, we have auxiliary information such as location (**LOC**, bivariate scaled longitude and latitude) and elevation (ELEV). At the sample level, information is also available for response variables. We will consider the following response variables:

- BIOMASS - total wood biomass per acre in tons;
- CRCOV - percent crown cover;
- BA - tree basal area per acre;
- NVOLTOT - total cuft volume per acre;
- FOREST - forest/nonforest indicator.

2.2 Variance estimation

To estimate the variance of systematic sample mean, we consider two design-based variance estimators, $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$, and the model-based variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$. The SRS variance estimator $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ is defined in (1.2). The stratified sampling variance estimator $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ is similar to the nonoverlapping differences estimator

$\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$ that we have discussed before. The difference is that for $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ we use 4-per-stratum design instead of 2-per-stratum for $\hat{\mathrm{V}}_{NO}(\bar{Y}_S)$. The expression for $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ is

$$\hat{\mathrm{V}}_{ST}(\bar{Y}_S) = \frac{1-f}{n}\frac{1}{n}\sum_{h=1}^{H}\frac{n_h}{n_h-1}\sum_{j\in S_h}(Y_j - \bar{Y}_{S_h})^2.$$

To construct $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$, we first divide the population into strata. Then we choose four points per stratum. Figure 2.2 displays this 4-per-stratum design. Note that for points near the edge of the map, there may be less than four points per stratum. We allow strata of size two or three, but if a stratum has only one sample point, we will combine it with its closest neighbor. It is possible that its neighbor has four points, so some strata may have five sample points.

For the purpose of constructing the model-based nonparametric variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, we consider the following model with location (**LOC**) as bivariate auxiliary variables:

$$Y_j = m(\mathbf{LOC}_j) + \varepsilon_j, \tag{2.1}$$

We assume errors to be independent and have homogeneous variance.

Under model (2.1), the nonparametric variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is

$$\hat{\mathrm{V}}_{NP}(\bar{Y}_S) = \frac{1}{kn^2}(\hat{\mathbf{m}}_b^T\mathbf{D}\hat{\mathbf{m}}_b) + \frac{1}{kn^2}\mathrm{tr}(\mathbf{D})\hat{\sigma}_{1b}^2,$$

where

$$\hat{\sigma}_{1b}^2 = \frac{1}{n}\sum_{j\in S_b}(Y_j - \hat{m}(\mathbf{LOC}_j))^2.$$

Here $m(\cdot)$ is estimated by bivariate local linear regression, and the estimator $\hat{m}(\cdot)$ is obtained using loess() in R. Since the samples points are approximated equally spaced (5×5 km grid), using loess() will produce very similar results to the fixed bandwidth local linear regression. We choose three spans: 0.1, 0.2 and 0.5. Figure 2.3 to Figure 2.5

display the contour plots of fitted response variable BIOMASS vs. **LOC**, with span = 0.5, 0.2 and 0.1, respectively.

We can see that Figure 2.5 has the right amount of smoothness. Figure 2.3 seems too smooth to capture enough details of the model. We also present the contour plot of other response variables with span = 0.1 in Figure 2.6 to Figure 2.9.

After obtaining $\hat{m}(\cdot)$, we can calculate the nonparametric variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ for each response variable. Table 2.1 presents the sample mean and the estimated variance of response variables using $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$. For $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, use model (2.1) and span = 0.5, 0.2 and 0.1, respectively.

	$\bar{Y}_S$	$\hat{\mathrm{V}}_{SRS}$	$\hat{\mathrm{V}}_{ST}$	$\hat{\mathrm{V}}_{NP0.5}$	$\hat{\mathrm{V}}_{NP0.2}$	$\hat{\mathrm{V}}_{NP0.1}$
BIOMASS	14.5	0.46	0.36	0.40	0.38	0.37
CRCOV	22.5	0.71	0.62	0.64	0.62	0.59
BA	48.5	3.87	3.19	3.40	3.30	3.12
NVOLTOT	906.9	1886	1538	1645	1584	1511
FOREST (%)	54.8	2.46	1.89	2.16	2.05	1.91

Table 2.1: Mean and variance estimates for five response variables. Five variance estimators are considered: $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ under model (2.1) with span = 0.5, 0.2 and 0.1, respectively.

From Table 2.1, we can see that for each response variable, $\hat{\mathrm{V}}_{SRS}$ always produces the biggest variance among the five variance estimators. The stratified variance estimator $\hat{\mathrm{V}}_{ST}$ is better than $\hat{\mathrm{V}}_{NP0.5}$ and $\hat{\mathrm{V}}_{NP0.2}$, and very close to $\hat{\mathrm{V}}_{NP0.1}$. The nonparametric variance estimator $\hat{\mathrm{V}}_{NP}$ gives smaller variance as span gets smaller. Note that in this real data problem where we do not know the true variance. However, we can consider the stratified variance estimator $\hat{\mathrm{V}}_{ST}$ to be a good choice because the 4-per-stratum design should be able to capture the population trend very well. We like the fact that $\hat{\mathrm{V}}_{NP}$ produces similar results to $\hat{\mathrm{V}}_{ST}$ if we choose a proper span, i.e. 0.1 in this example. However, the nonparametric variance estimator $\hat{\mathrm{V}}_{NP}$ can take the advantage of using a model. For example, if we are to put more model variables in the model, we may even further improve our results.

So we consider a more sophisticated model as follows, which also includes elevation (ELEV) in additive to **LOC**:

$$Y_j = m_1(\mathbf{LOC}_j) + m_2(ELEV_j) + \varepsilon_j. \tag{2.2}$$

We fit model (2.2) in R using Generalized Additive Models (gam) package. To choose the span, we first consider the same span for both **LOC** and ELEV. Figure 2.10 displays the contour plot of fitted response variable BIOMASS vs. **LOC**, where span = 0.1 for both **LOC** and ELEV.

Table 2.2 reports the sample mean and the estimated variance of response variables using $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ (same as those in Table 2.1). For $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$, use model (2.2) and span = 0.5, 0.2 and 0.1, respectively. Same span is used for both **LOC** and ELEV.

	$\bar{Y}_S$	$\hat{\mathrm{V}}_{SRS}$	$\hat{\mathrm{V}}_{ST}$	$\hat{\mathrm{V}}_{NP0.5}$	$\hat{\mathrm{V}}_{NP0.2}$	$\hat{\mathrm{V}}_{NP0.1}$
BIOMASS	14.5	0.46	0.36	0.36	0.34	0.33
CRCOV	22.5	0.71	0.62	0.59	0.55	0.53
BA	48.5	3.87	3.19	3.11	2.96	2.78
NVOLTOT	906.9	1886	1538	1487	1417	1342
FOREST (%)	54.8	2.46	1.89	1.92	1.77	1.65

Table 2.2: Mean and variance estimates for five response variables. Five variance estimators are considered: $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ under model (2.2) with span = 0.5, 0.2 and 0.1, respectively. Same span is used for both **LOC** and ELEV.

We can see that, if we add model variable ELEV, we further decrease $\hat{\mathrm{V}}_{NP}$. Now the nonparametric estimator $\hat{\mathrm{V}}_{NP}$ with even the biggest span 0.5 has the same or smaller estimates than $\hat{\mathrm{V}}_{ST}$ for all response variables.

Now let us examine if elevation (ELEV) is indeed a useful predictor variable. Figure 2.11 displays the contour plot of ELEV vs. **LOC**.

As we can see, there is an obvious spatial trend of elevation in this study region, suggesting that elevation may play a roll in the values of response variables. On a more detailed level, we calculate the fitted response variable contributed by elevation, i.e. $\hat{m}_2(ELEV_j)$, centered around 0, and plot it against elevation. Take BIOMASS as an example, this plot is displayed in Figure 2.12.

We can see that, caused by elevation, BIOMASS increases until elevation is around 3200 meters, then decreases as elevation gets higher. This trend, combined with Figure 2.11, suggest that elevation (ELEV) is a useful predictor variable. We also calculate the fitted BIOMASS contributed by location, i.e. $\hat{m}_1(\mathbf{LOC}_j)$, centered around 0, and produce the contour plot against location. This contour plot is displayed in Figure 2.13.

We can see that in Figure 2.12, the curve is a bit too wiggly, suggesting that span = 0.1 may be too small for elevation. In Figure 2.13, span = 0.1 seems a reasonable choice. So instead of using the same span for both **LOC** and ELEV, we use span = 0.1 for **LOC** and 0.3 for ELEV. The new gam component plots for BIOMASS are presented in Figure 2.14.

Now the smoothness of both components look good. We produce the same gam component plots for the other four response variables. The plots are shown in Figure 2.15 to Figure 2.18.

Let $\hat{\mathrm{V}}_{NP(0.1,\,0.3)}$ denote the nonparametric variance estimator $\hat{\mathrm{V}}_{NP}$ using span = 0.1 for location and span = 0.3 for elevation. We calculate the values of $\hat{\mathrm{V}}_{NP(0.1,\,0.3)}$ and attach the results to Table 2.2. The updated table is Table 2.3.

	$\bar{Y}_S$	$\hat{\mathrm{V}}_{SRS}$	$\hat{\mathrm{V}}_{ST}$	$\hat{\mathrm{V}}_{NP0.5}$	$\hat{\mathrm{V}}_{NP0.2}$	$\hat{\mathrm{V}}_{NP0.1}$	$\hat{\mathrm{V}}_{NP(0.1,\,0.3)}$
BIOMASS	14.5	0.46	0.36	0.36	0.34	0.33	0.34
CRCOV	22.5	0.71	0.62	0.59	0.55	0.53	0.55
BA	48.5	3.87	3.19	3.11	2.96	2.78	2.87
NVOLTOT	906.9	1886	1538	1487	1417	1342	1396
FOREST (%)	54.8	2.46	1.89	1.92	1.77	1.65	1.71

Table 2.3: Mean and variance estimates for five response variables. Six variance estimators are considered: $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$, $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ under model (2.2) with span = 0.5, 0.2 and 0.1, respectively. Same span is used for both **LOC** and ELEV. And $\hat{\mathrm{V}}_{NP(0.1,\,0.3)}$ where span = 0.1 for location and span = 0.3 for elevation.

We can see that $\hat{\mathrm{V}}_{NP(0.1,\,0.3)}$ is generally larger than $\hat{\mathrm{V}}_{NP0.1}$, but still smaller than $\hat{\mathrm{V}}_{ST}$.

2.3 Conclusions

Based on our discussions in previous sections, we come to the following conclusions:

1. The nonparametric variance estimator $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ produces very good estimates for the variance in this FIA example. The results are close to $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$, which we believe to be a good estimator because the stratification should capture the spatial trend very well. Both $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ and $\hat{\mathrm{V}}_{ST}(\bar{Y}_S)$ are better than $\hat{\mathrm{V}}_{SRS}(\bar{Y}_S)$.

2. An advantage of using $\hat{\mathrm{V}}_{NP}(\bar{Y}_S)$ is the flexibility. We can include more auxiliary variables or change the bandwidth (span) of nonparametric fitting, and further improve the results.

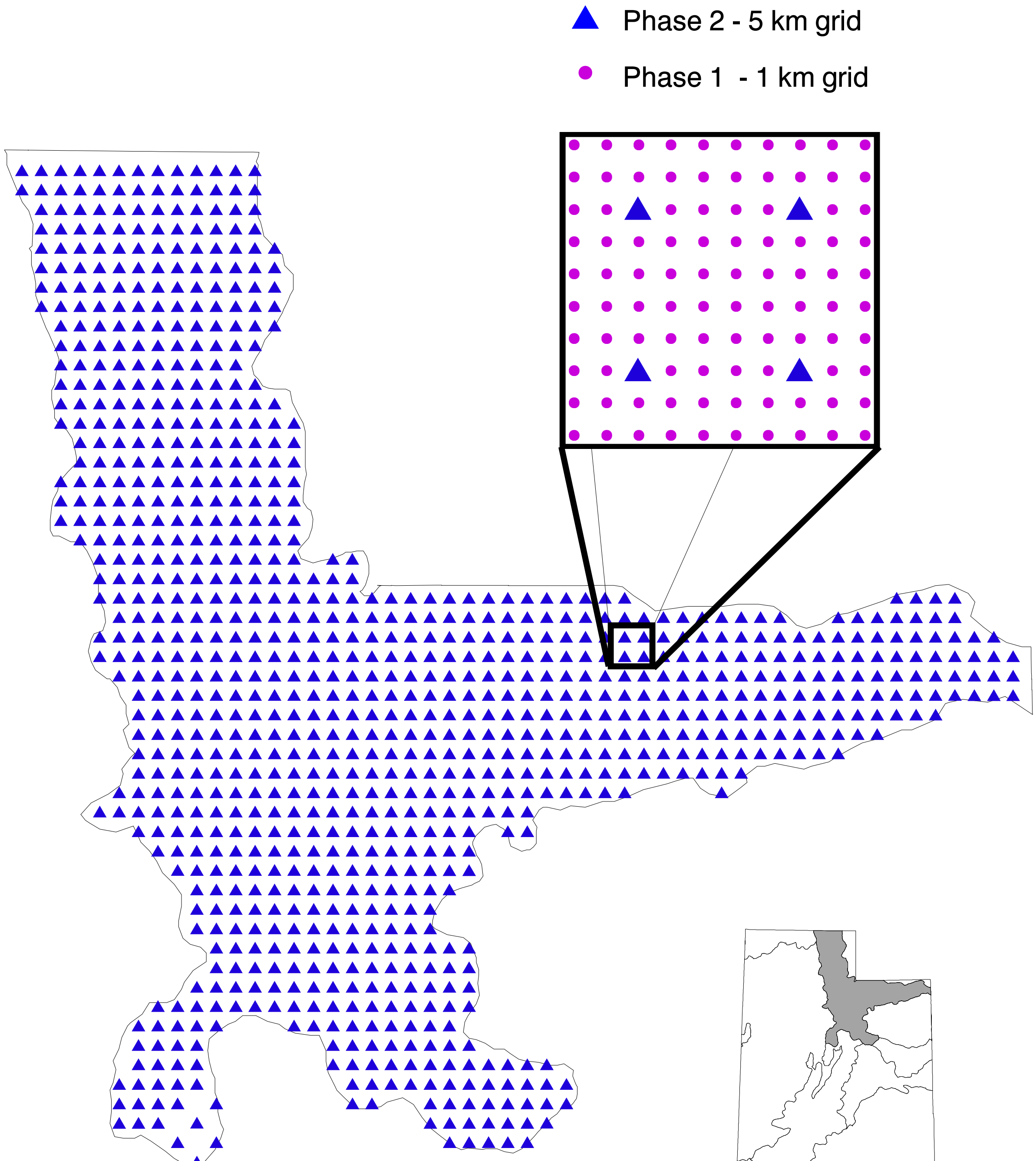


Figure 2.1: Map of the study region in northern Utah. Each triangle represents a field-visited phase two sample point. Each dot in the magnified section represent a phase one sample point. There are 24,980 phase one sample points and 968 phase two sample points.

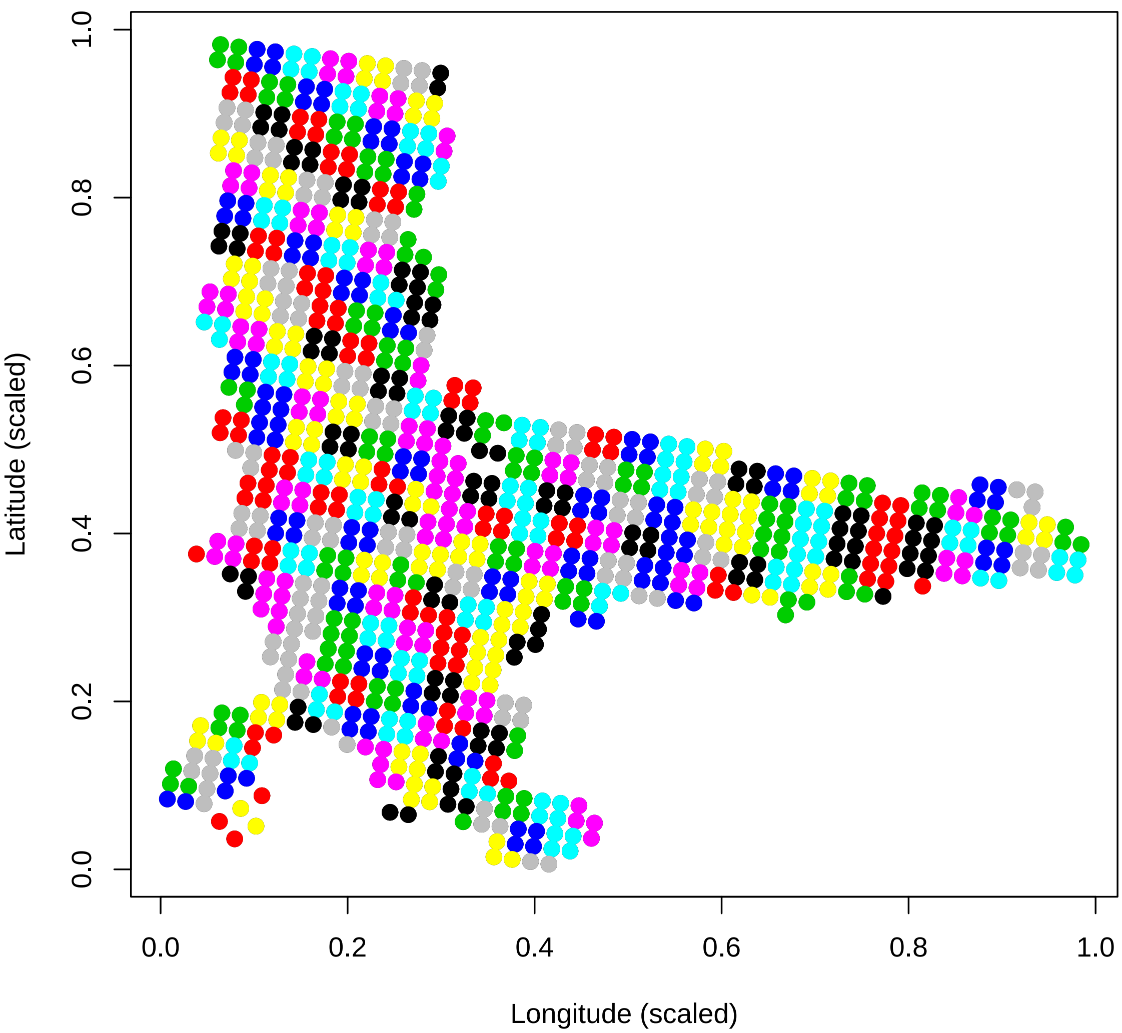


Figure 2.2: Illustration of 4-per-stratum design. Each color/shade represents a stratum. For points near the edge of the map, there may be less than four points per stratum. We allow strata of size two or three, but if there is one point in a stratum, we will merge that point to its closest neighbor stratum.

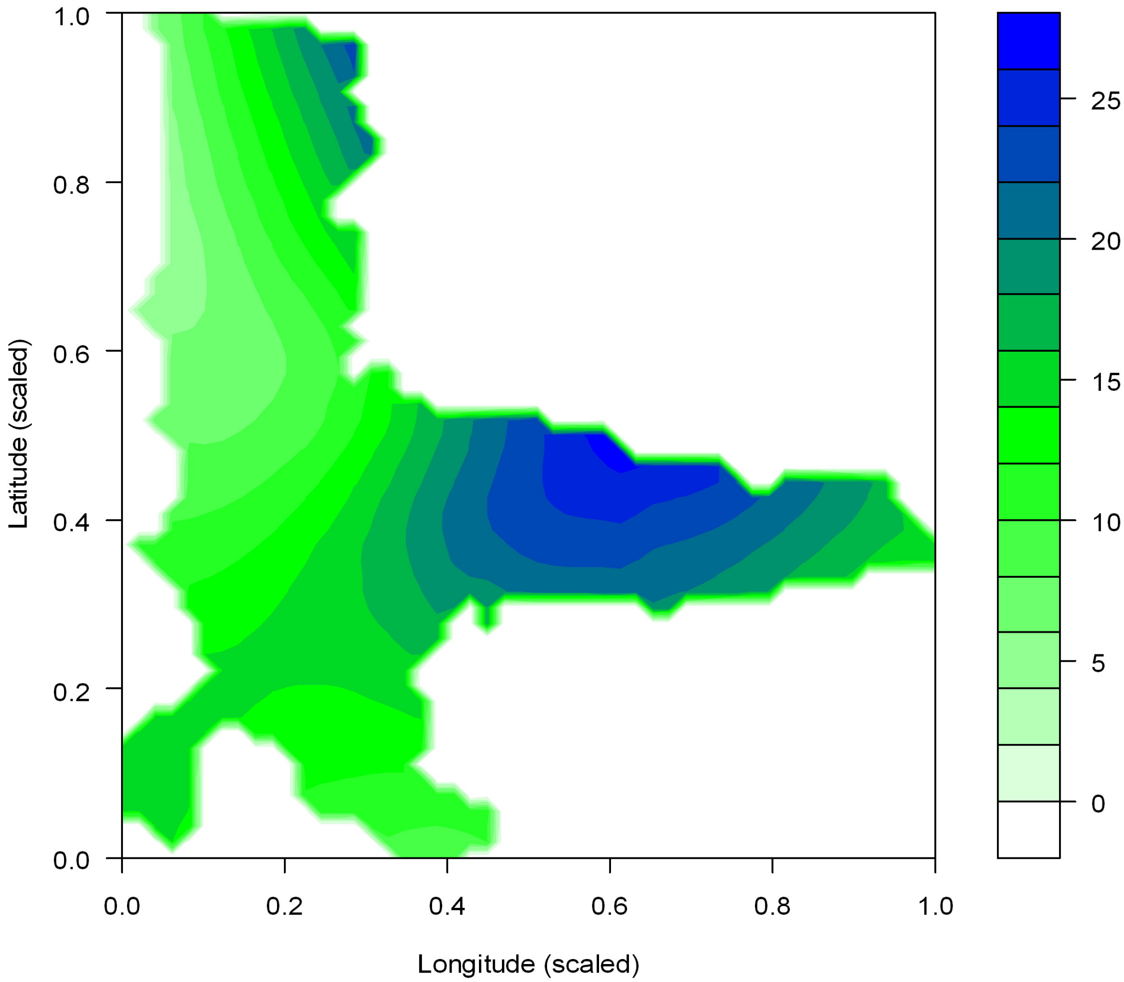


Figure 2.3: Contour plot of fitted BIOMASS vs. location (**LOC**) in the study region. The model function is estimated using loess() in R with span 0.5.

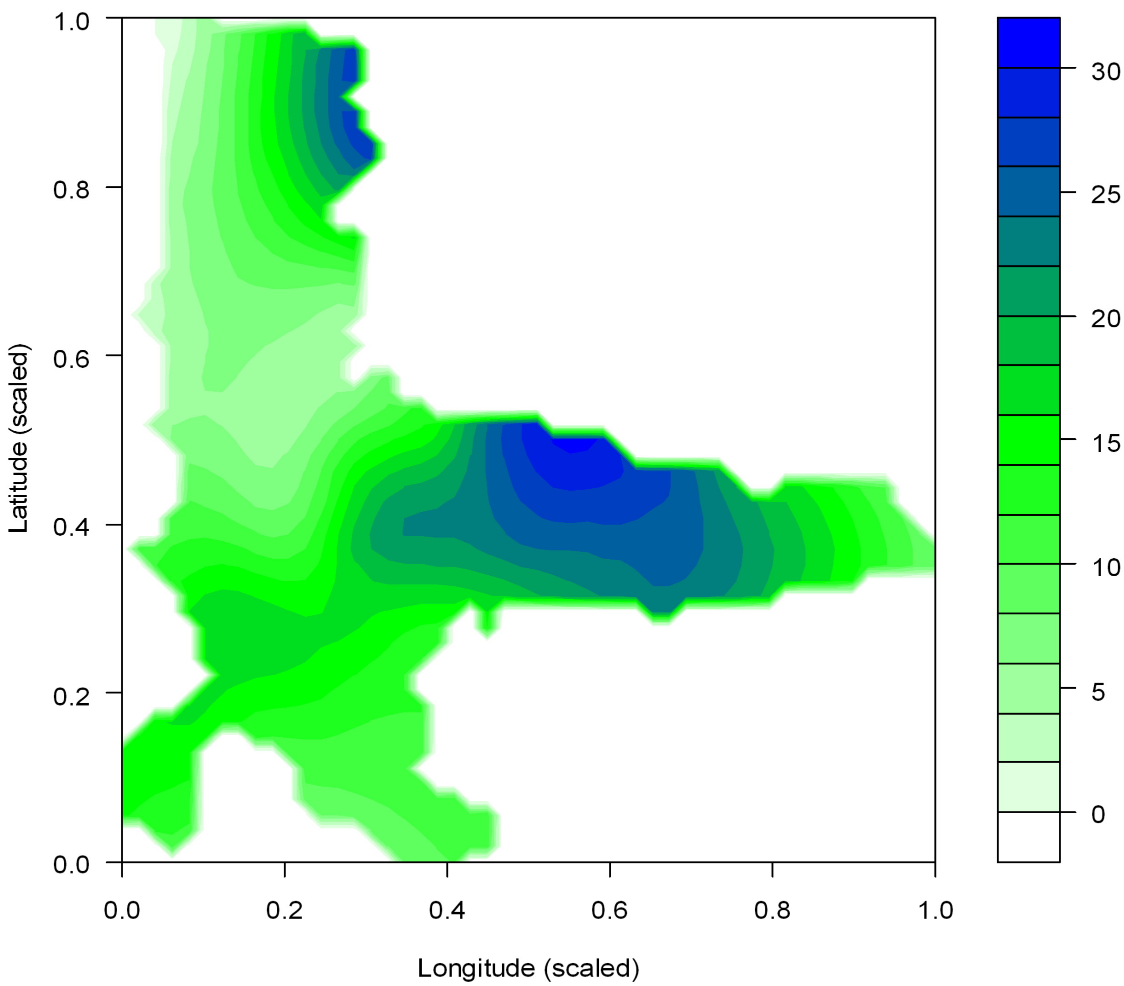


Figure 2.4: Contour plot of fitted BIOMASS vs. location (**LOC**) in the study region. The model function is estimated using loess() in R with span 0.2.

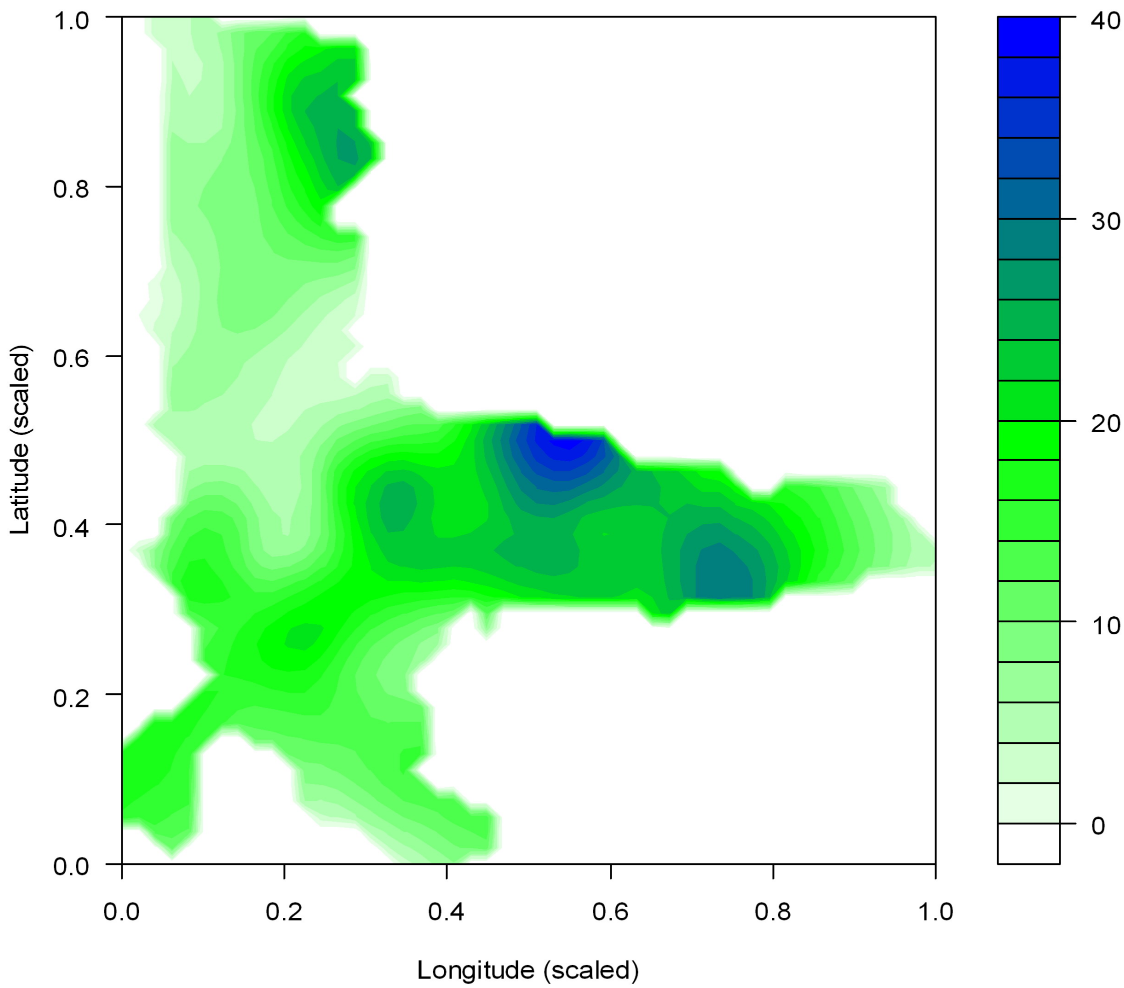


Figure 2.5: Contour plot of fitted BIOMASS vs. location (**LOC**) in the study region. The model function in (2.1) is estimated using loess() in R with span 0.1.

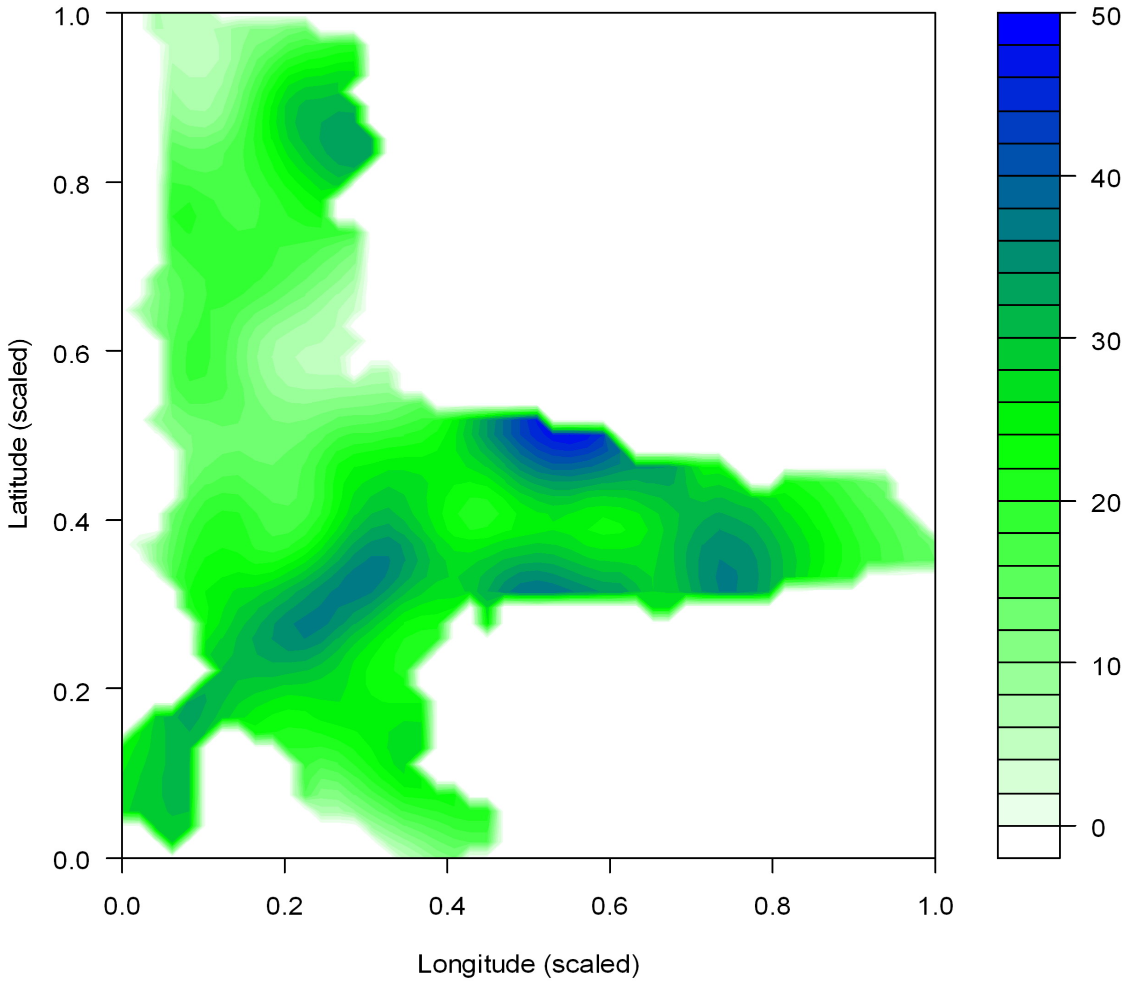


Figure 2.6: Contour plot of fitted CRCOV vs. location (**LOC**) in the study region. The model function in (2.1) is estimated using loess() in R with span 0.1.

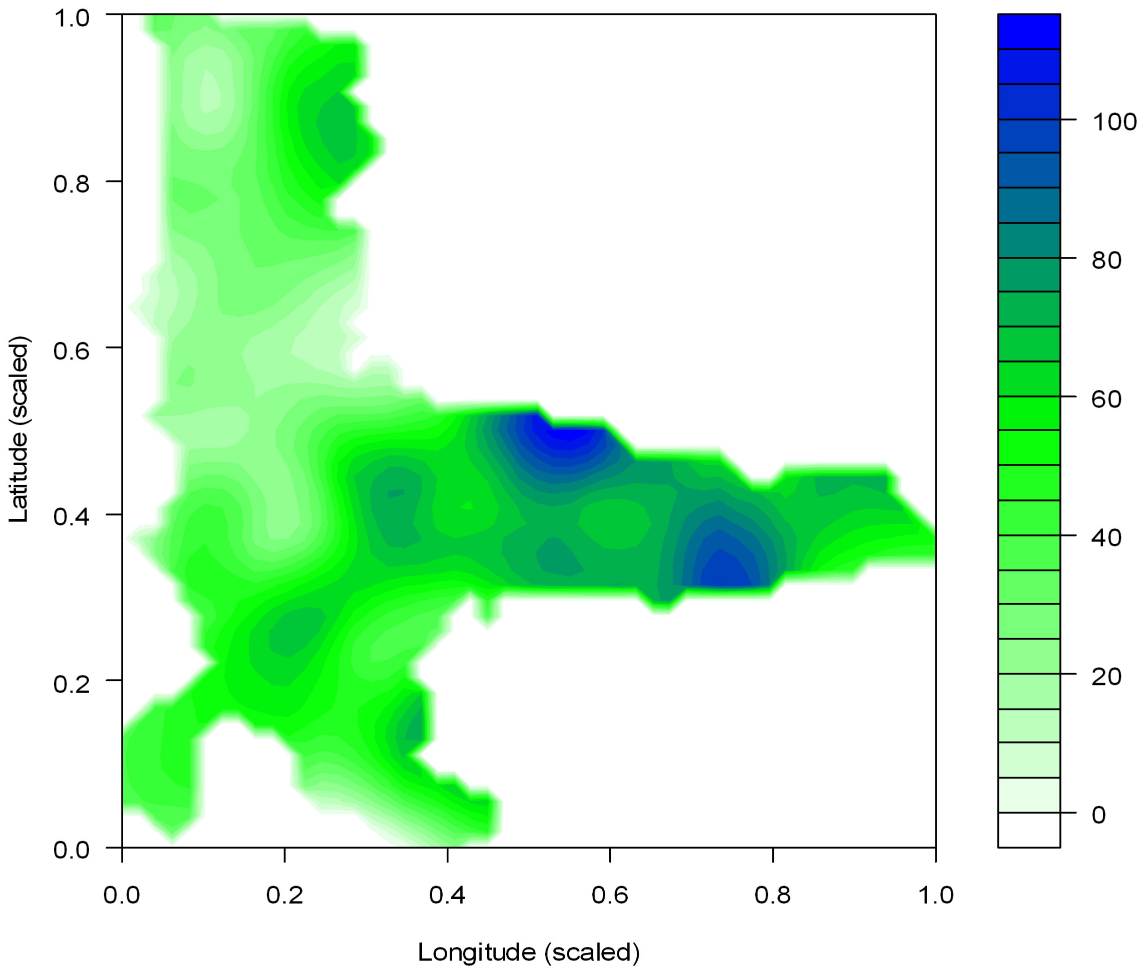


Figure 2.7: Contour plot of fitted BA vs. location (**LOC**) in the study region. The model function in (2.1) is estimated using loess() in R with span 0.1.

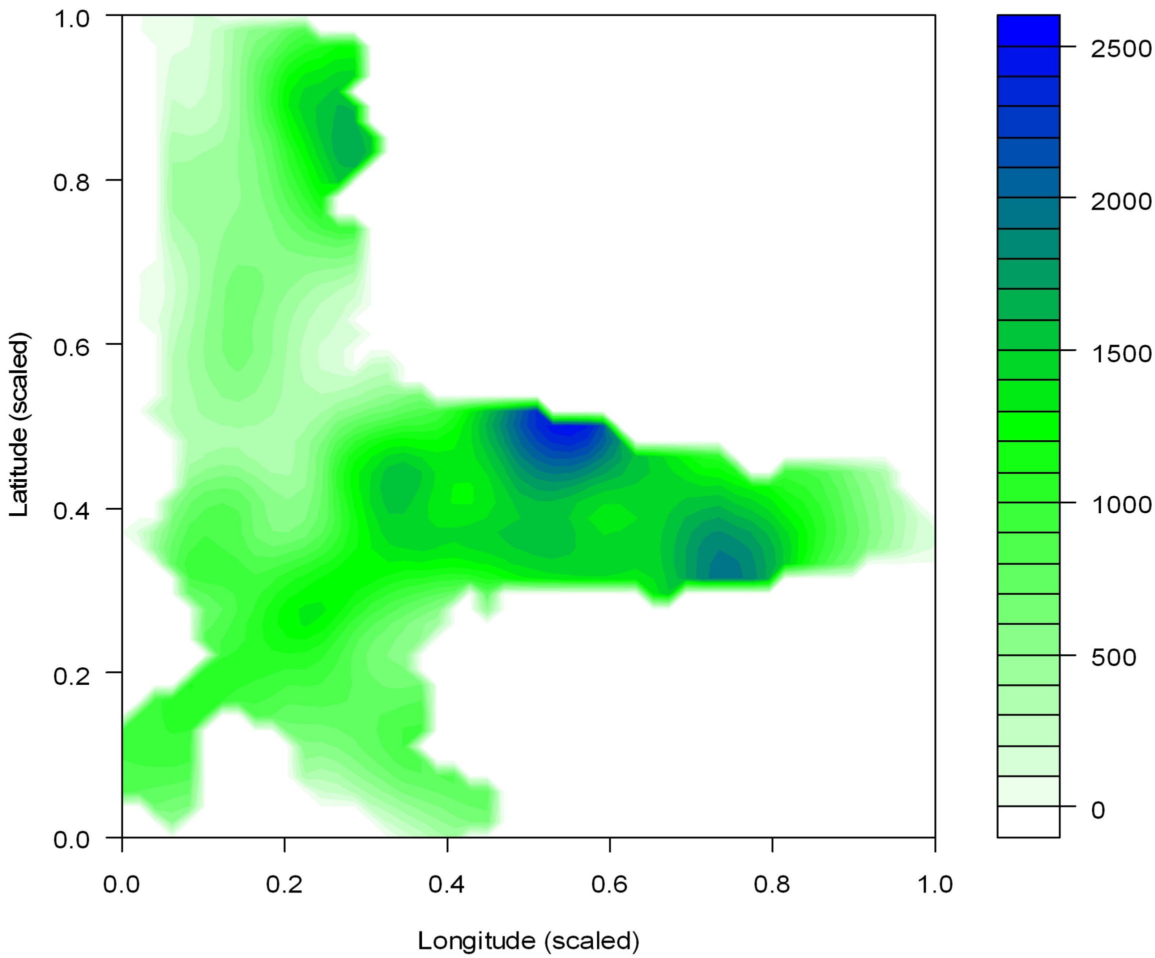


Figure 2.8: Contour plot of fitted NVOLTOT vs. location (**LOC**) in the study region. The model function in (2.1) is estimated using loess() in R with span 0.1.

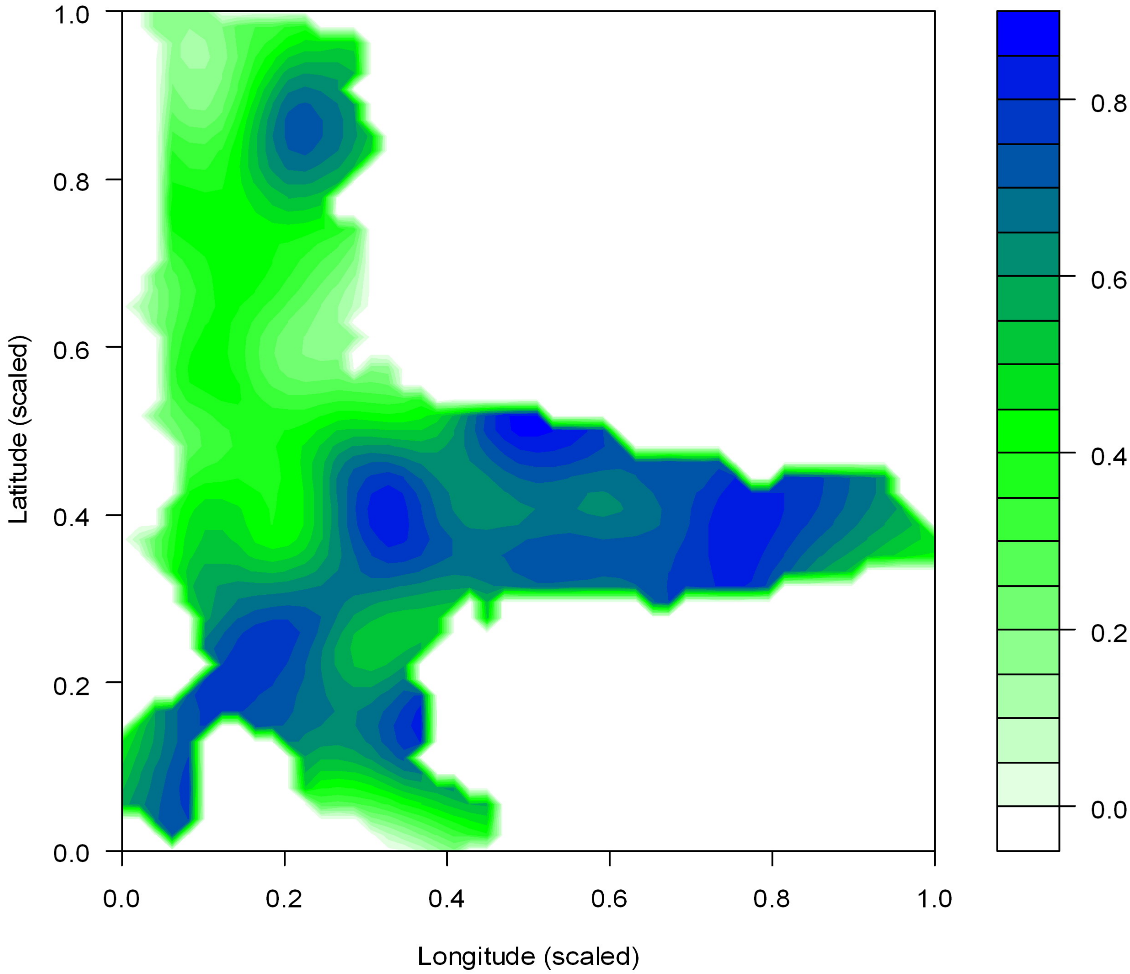


Figure 2.9: Contour plot of fitted FOREST vs. location (**LOC**) in the study region. The model function in (2.1) is estimated using loess() in R with span 0.1.

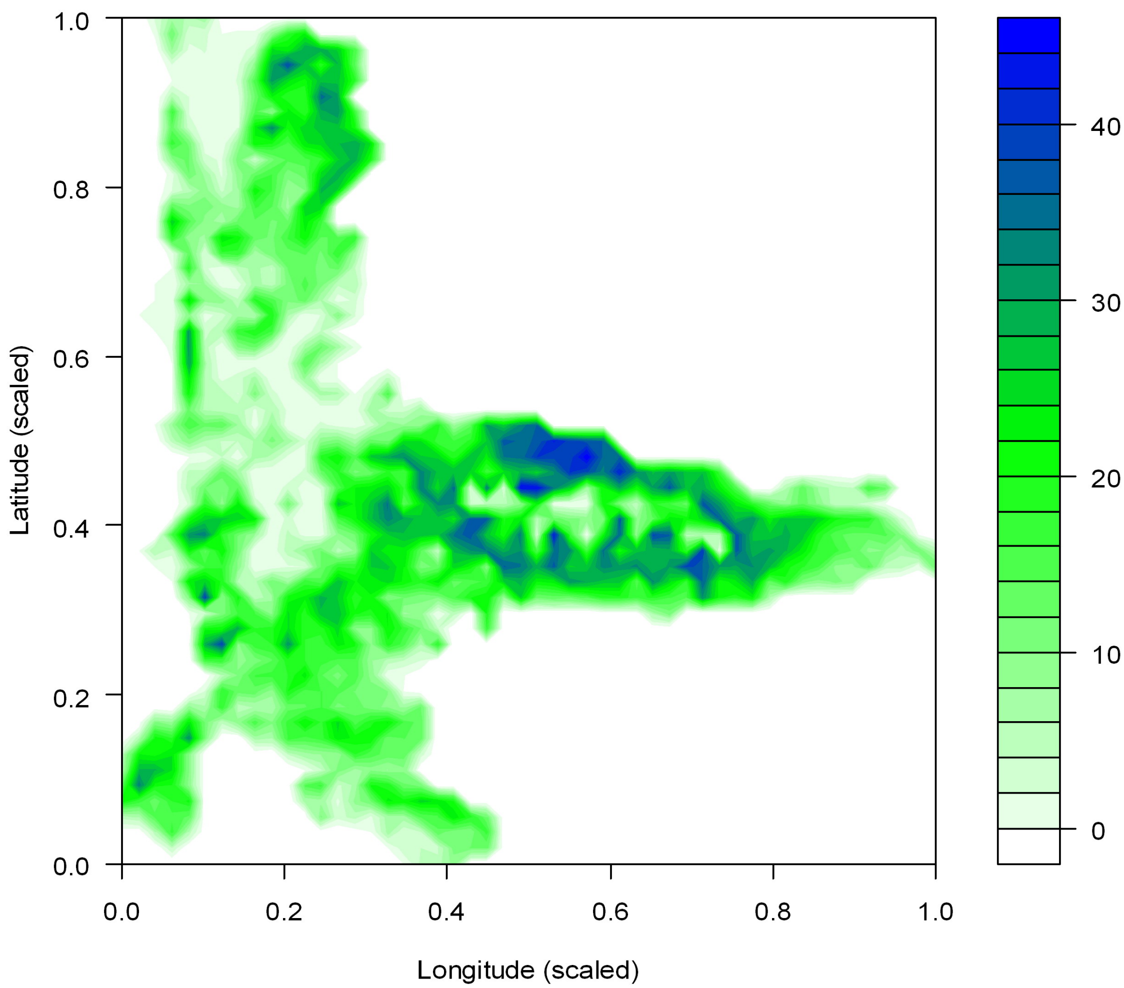


Figure 2.10: Contour plot of fitted BIOMASS vs. location (**LOC**) in the study region. The model function in (2.2) is estimated using gam() in R with span 0.1 for both **LOC** and ELEV.

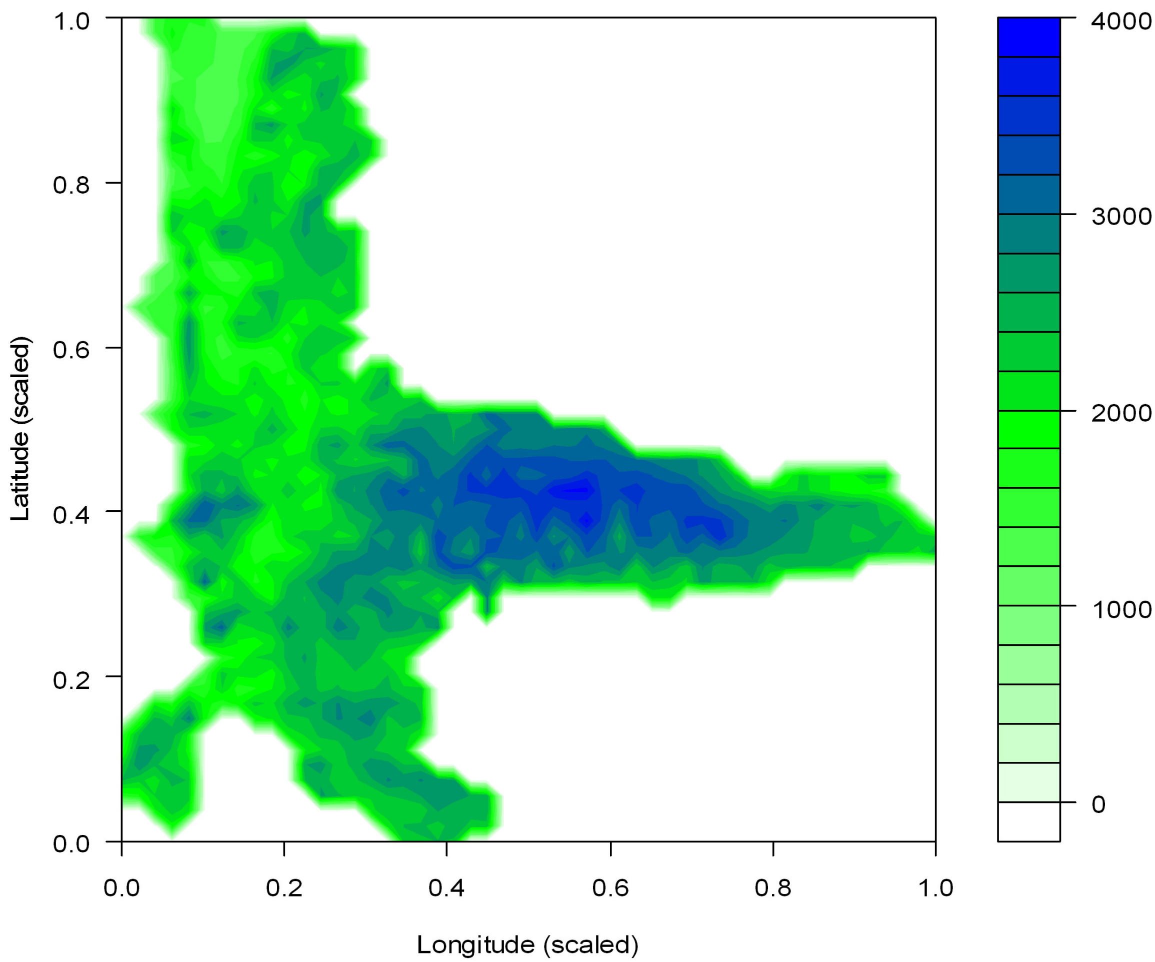


Figure 2.11: Contour plot of elevation vs. location.

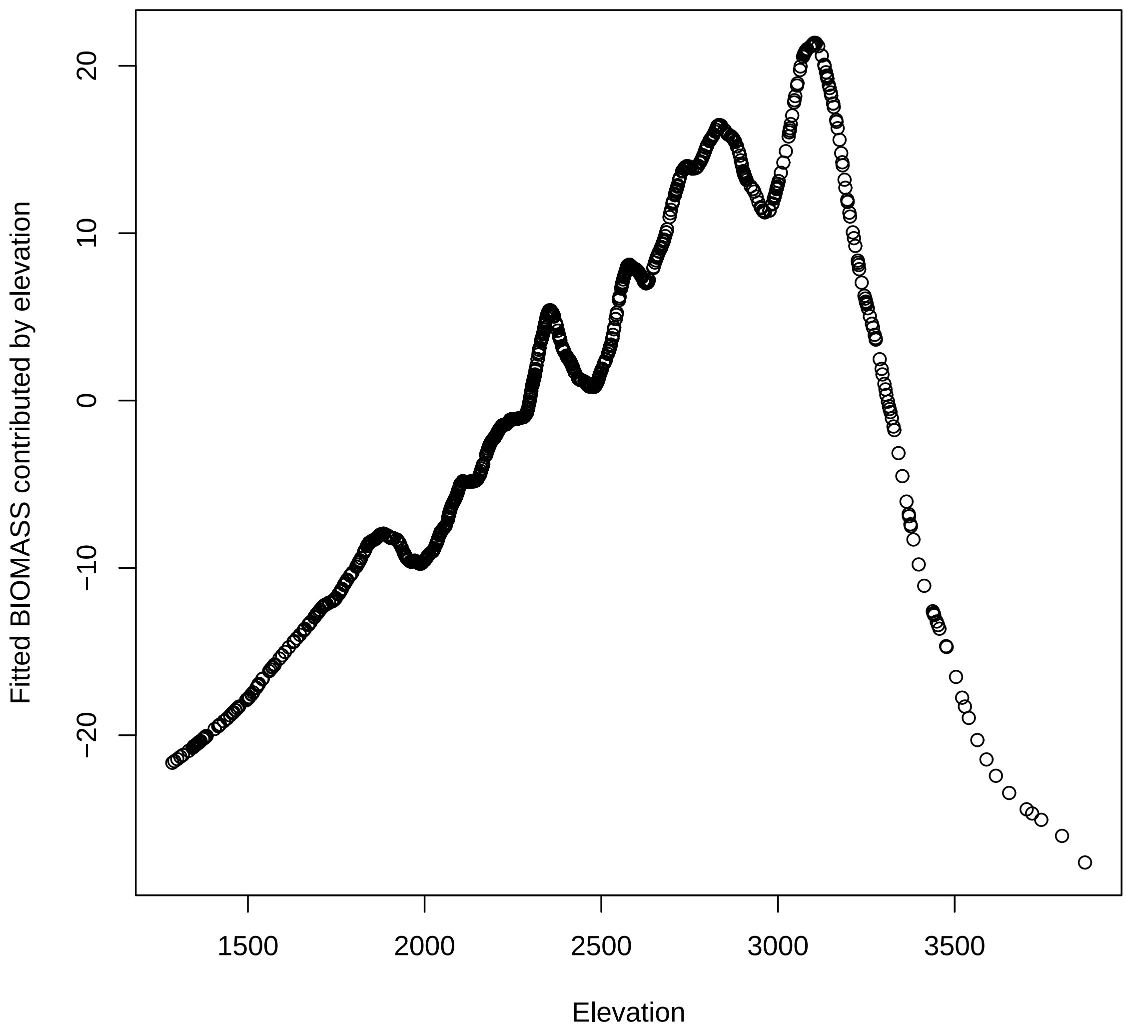


Figure 2.12: Fitted BIOMASS contributed by elevation, centered around 0, against elevation. Same span = 0.1 is used for both **LOC** and ELEV.

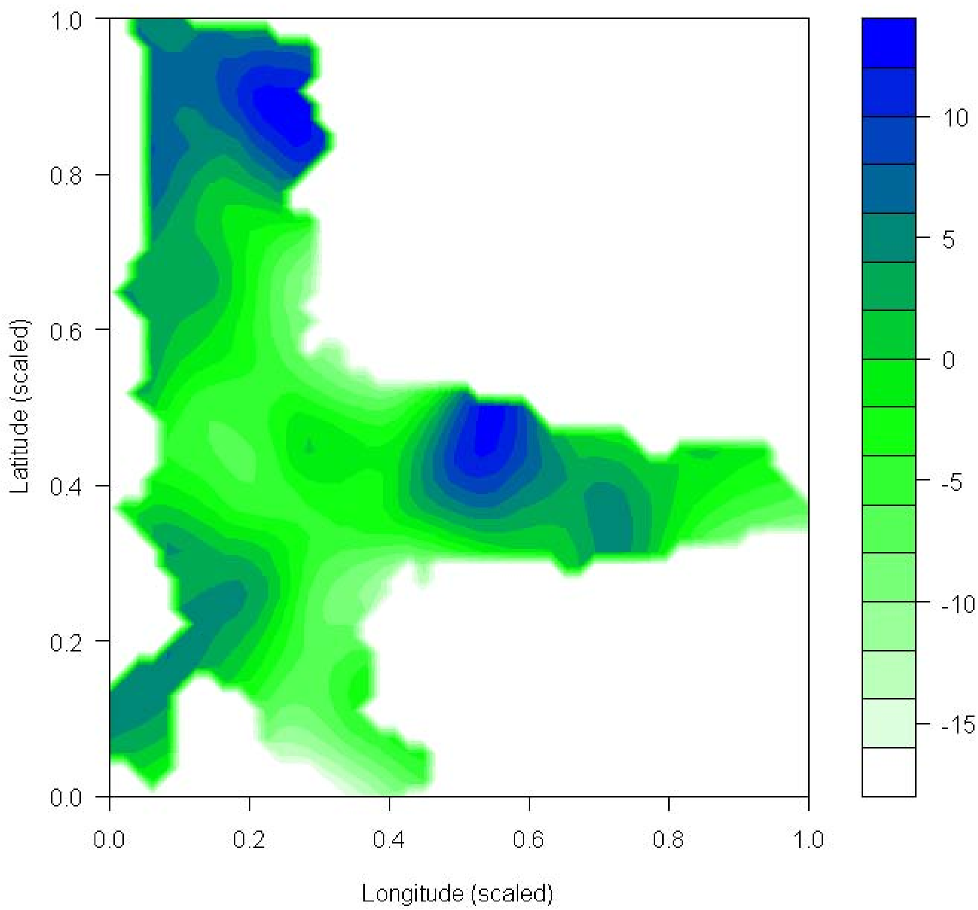


Figure 2.13: Contour plot of fitted BIOMASS contributed by location, centered around 0, against location. Same span = 0.1 is used for both **LOC** and ELEV.

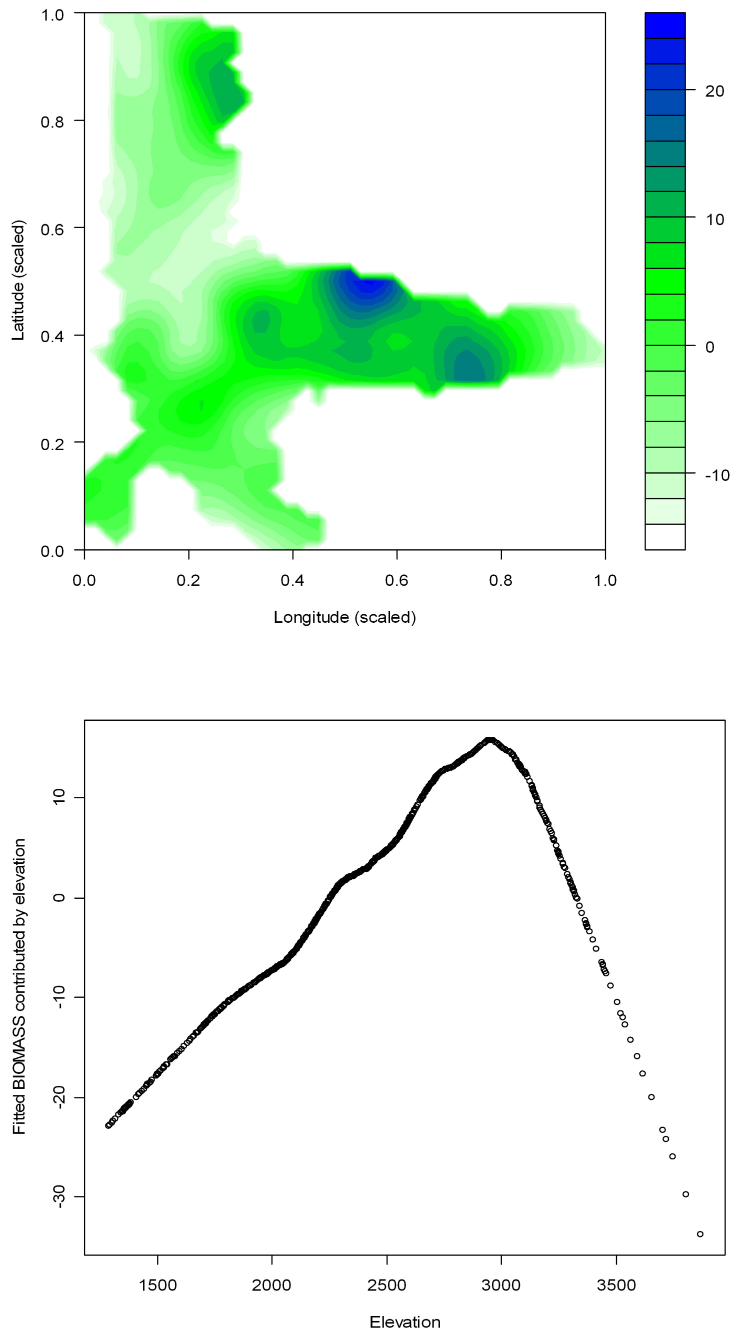
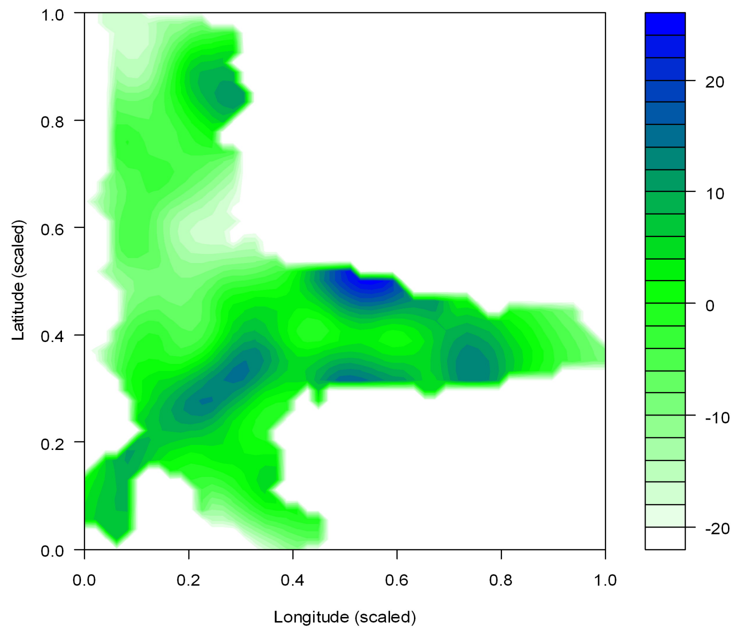


Figure 2.14: Gam component (centered around 0) plots for response variable BIOMASS. Top plot is centered $\hat{m}_1(\mathbf{LOC}_j)$ vs **LOC** and bottom one is centered $\hat{m}_1(ELEV_j)$ vs ELEV. Span = 0.1 is used for **LOC** and span = 0.3 is used for ELEV.



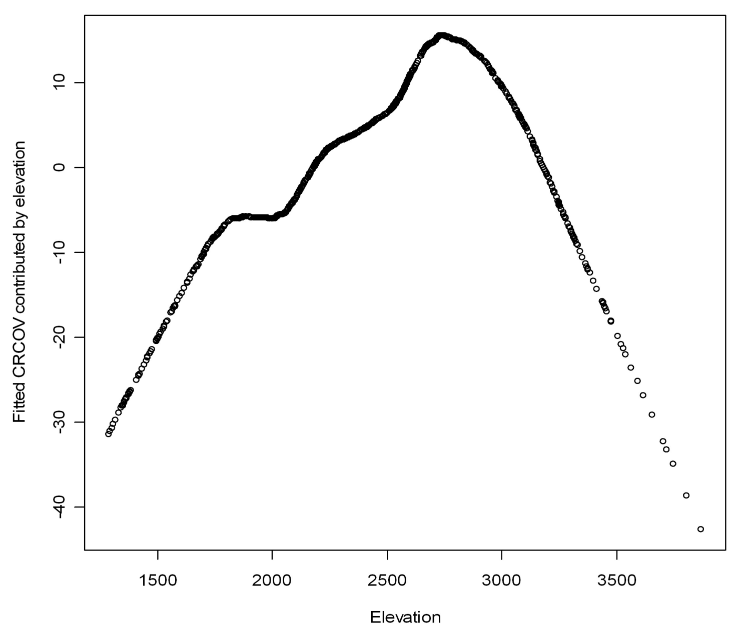


Figure 2.15: Gam component (centered around 0) plots for response variable CRCOV. Top plot is centered $\hat{m}_1(\mathbf{LOC}_j)$ vs **LOC** and bottom one is centered $\hat{m}_1(ELEV_j)$ vs ELEV. Span = 0.1 is used for **LOC** and span = 0.3 is used for ELEV.

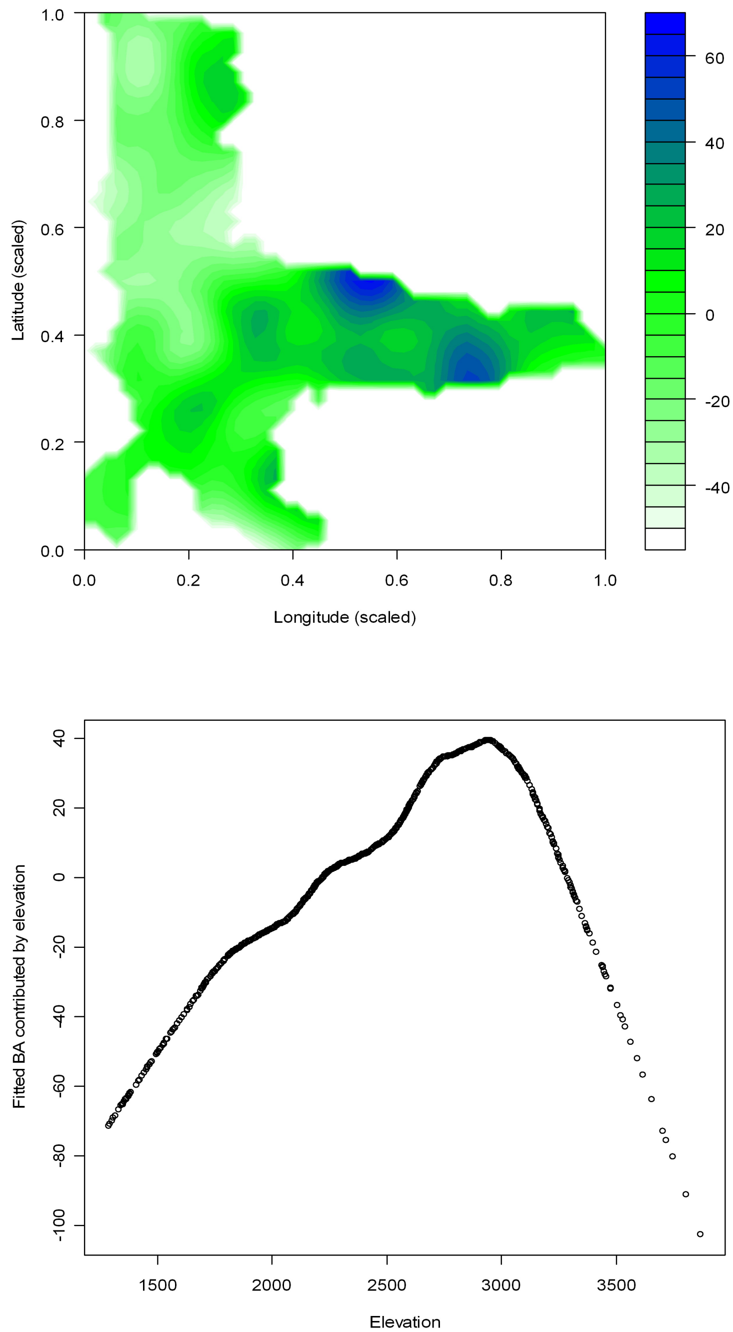
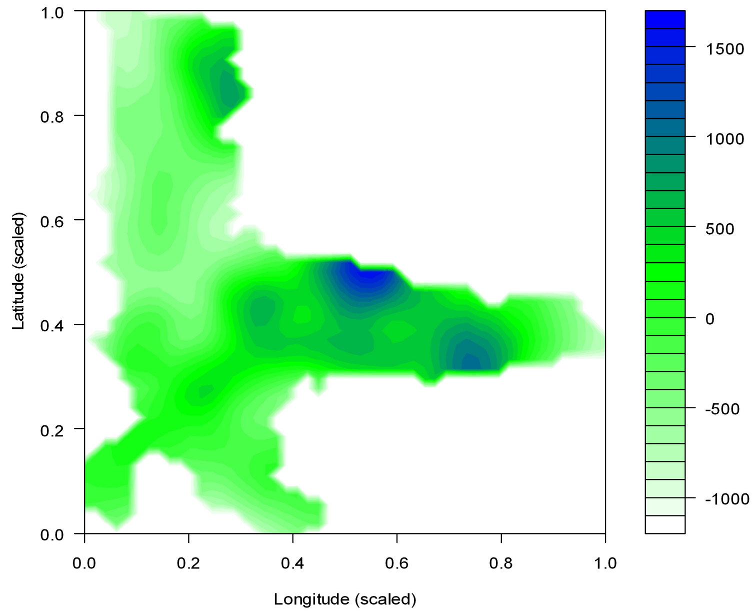


Figure 2.16: Gam component (centered around 0) plots for response variable BA. Top plot is centered $\hat{m}_1(\mathbf{LOC}_j)$ vs **LOC** and bottom one is centered $\hat{m}_1(ELEV_j)$ vs ELEV. Span = 0.1 is used for **LOC** and span = 0.3 is used for ELEV.



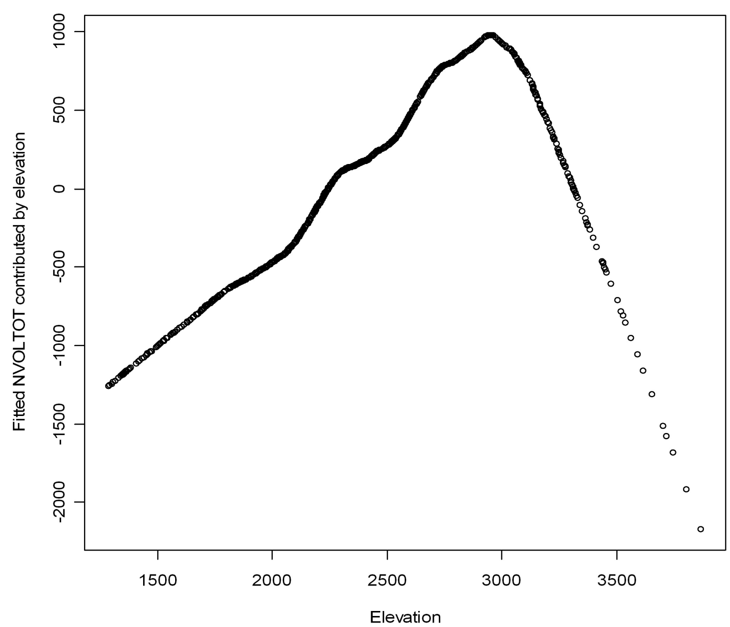


Figure 2.17: Gam component (centered around 0) plots for response variable NVOLTOT. Top plot is centered $\hat{m}_1(\mathbf{LOC}_j)$ vs **LOC** and bottom one is centered $\hat{m}_1(ELEV_j)$ vs ELEV. Span = 0.1 is used for **LOC** and span = 0.3 is used for ELEV.

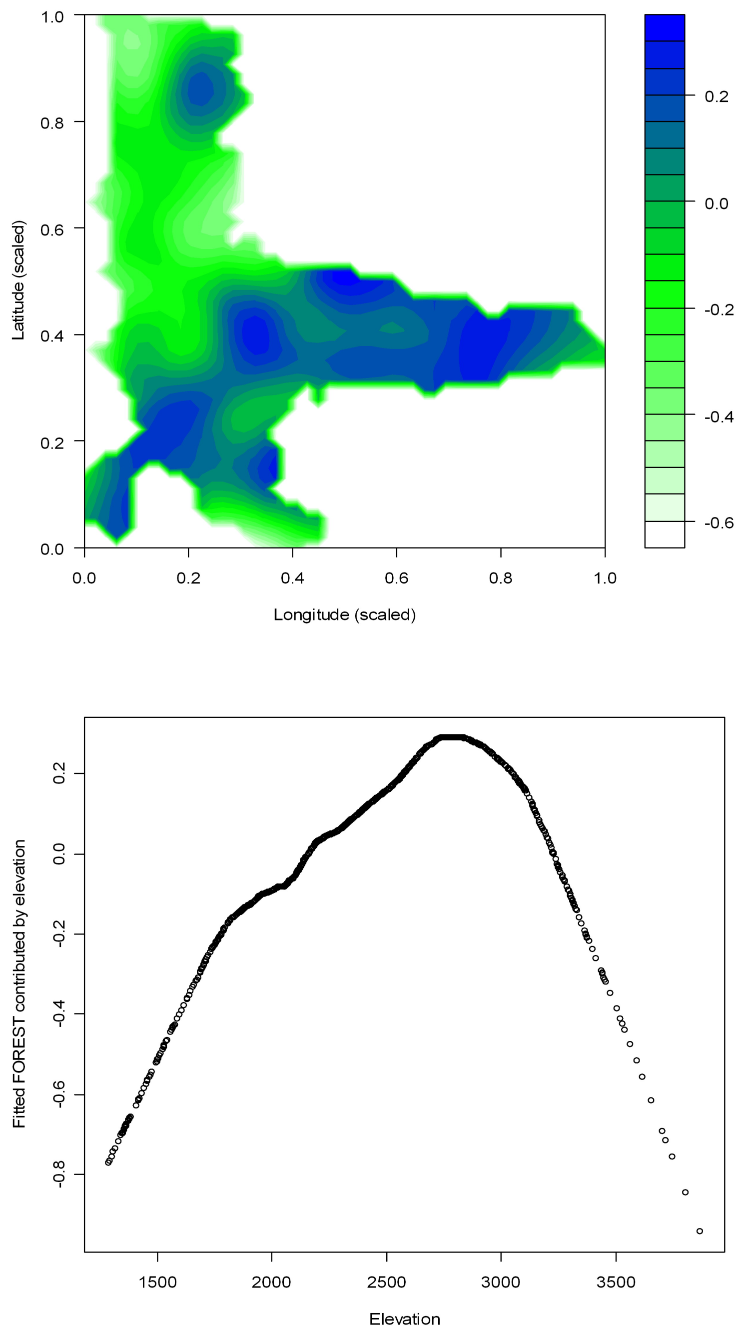


Figure 2.18: Gam component (centered around 0) plots for response variable FOREST. Top plot is centered $\hat{m}_1(\mathbf{LOC}_j)$ vs **LOC** and bottom one is centered $\hat{m}_1(ELEV_j)$ vs ELEV. Span = 0.1 is used for **LOC** and span = 0.3 is used for ELEV.

CHAPTER 3 Model averaging in survey estimation

3.1 Regression estimator

Regression estimator is often used in survey estimation. It makes use of the auxiliary information about the population to provide efficient estimation. Suppose we have l study variables $\mathbf{Y}_j \in \Re^l$ and p auxiliary variables $\mathbf{X}_j \in \Re^p$. Both $\mathbf{Y}_j$ and $\mathbf{X}_j$ are row vectors. Let Y_j denote the jth element of one of the study variables. Consider the following linear regression model

$$Y_j = \mathbf{X}_j \boldsymbol{\beta} + \varepsilon_j,$$

where ε_j's are independent random variables with mean zero and variance $\sigma^2 \omega_j$. We observe the study variable Y_j for $j \in S$, and the auxiliary variables $\mathbf{X}_j$ for $j \in U$. The population U is of size N and the sample S is of size n. Suppose the quantity of interest is the population total $t_y = \sum_{j \in U} Y_j$. Särndal et al. (1992) proposed a regression estimator for the population total of the form

$$\hat{t}_{reg} = \sum_{j \in S} \frac{Y_j - \hat{Y}_j}{\pi_j} + \sum_{j \in U} \hat{Y}_j,$$

where $\hat{Y}_j = \mathbf{X}_j \hat{\boldsymbol{\beta}}_S$ is the predicted value for Y_j and $\hat{\boldsymbol{\beta}}_S$ is the weighted least square (WLS) estimator for $\boldsymbol{\beta}$ obtained from the sample S. Specifically,

$$\hat{\boldsymbol{\beta}}_S = (\mathbf{X}_S^T \boldsymbol{\Omega}_S^{-1} \mathbf{X}_S^T)^{-1} \mathbf{X}_S^T \boldsymbol{\Omega}_S^{-1} \mathbf{Y}_S,$$

where $\boldsymbol{\Omega}_S = \text{diag}\{\omega_j\}$, $j \in S$. We assume that ω_j's are known up to constants. This form shows that the regression estimator is a sum of population total of fitted values,

$\sum_{j\in U} \hat{Y}_j$, and an adjustment term $\sum_{j\in S}(Y_j - \hat{Y}_j)/\pi_j$.

The efficiency of regression estimator is measured by its variance. Using Taylor linearization, Särndal et al. (1992) showed that the approximate variance for $\hat{t}_{reg}$ is

$$AV(\hat{t}_{reg}) = \sum_{j\in U}\sum_{i\in U}(\pi_{ji} - \pi_j\pi_i)\frac{(Y_j - \mathbf{X}_j\mathbf{B})}{\pi_j}\frac{(Y_i - \mathbf{X}_i\mathbf{B})}{\pi_i},$$

where $\mathbf{B}$ is defined as

$$\mathbf{B} = (\mathbf{X}_U^T\mathbf{\Omega}_U^{-1}\mathbf{X}_U^T)^{-1}\mathbf{X}_U^T\mathbf{\Omega}_U^{-1}\mathbf{Y}_U,$$

with $\mathbf{\Omega}_U = \text{diag}\{\omega_j\}$, $j \in U$.

The variance estimator for $\hat{t}_{reg}$ is

$$\hat{V}(\hat{t}_{reg}) = \sum_{j\in S}\sum_{i\in S}\frac{\pi_{ji} - \pi_j\pi_i}{\pi_{ji}}\frac{(Y_j - \hat{Y}_j)}{\pi_j}\frac{(Y_i - \hat{Y}_i)}{\pi_i},$$

where π_j is the inclusion probability for the jth element in sample S and π_{ji} is the probability of including both the jth and the ith element in sample S.

The above discussion deals with a parametric approach for regression estimation. There are also nonparametric approaches. Let us consider the following model

$$Y_j = m(\mathbf{X}_j) + \varepsilon_j,$$

where m is a continuous and bounded function and ε_j's are independent random variables with mean zero and variance $\sigma^2\omega_j$. Let $\hat{m}_j$ denote the predicted model function for $m(\mathbf{x}_j)$ using nonparametric regression. Breidt and Opsomer (2000) proposed a model-assisted local polynomial regression estimator of the form

$$\hat{t}_y = \sum_{j\in S}\frac{Y_j - \hat{m}_j}{\pi_j} + \sum_{j\in U}\hat{m}_j. \tag{3.1}$$

For a single auxiliary variable X_j, given that $X_j = x_j$,

$$\begin{aligned} \hat{m}_j &= \mathbf{e}_1^T\left(\mathbf{X}_{Sj}^T\mathbf{W}_{Sj}\mathbf{X}_{Sj} + \frac{\nu}{N^{q+1}}\mathbf{I}\right)^{-1}\mathbf{X}_{Sj}^T\mathbf{W}_{Sj}\mathbf{Y}_S \qquad (3.2)\\ &= \mathbf{w}_{Sj}^T\mathbf{Y}_S, \end{aligned}$$

where q is the degrees of local polynomial regression, $\mathbf{e}_1$ is the $(q+1)\times 1$ vector having 1 in the first entry and all other entries 0, and

$$\mathbf{X}_{Sj} = \begin{pmatrix} 1 & (x_1 - x_j) & \cdots & (x_1 - x_j)^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_n - x_j) & \cdots & (x_n - x_j)^q \end{pmatrix},$$

$$\mathbf{W}_{Sj} = \operatorname{diag}\left\{ K\left(\frac{x_i - x_j}{h}\right)\frac{1}{\pi_j h}, \quad i \in S \right\},$$

where h is the bandwidth and $K\left(\frac{x_i - x_j}{h}\right)$ is the kernel function. On the right hand side of expression (3.2), the adjustment term $\frac{\nu}{N^{q+1}}\mathbf{I}$, where $\nu > 0$, is used to ensure the estimator $\hat{m}_j$ is well defined for all $S \subset U$.

Breidt and Opsomer (2000) showed that the asymptotic MSE of the local polynomial regression estimator $\hat{t}_y$ is equivalent to the variance of the generalized difference estimator, which is

$$\operatorname{Var}_p(t_y^*) = \sum_{j\in U}\sum_{i\in U}(\pi_{ji} - \pi_j\pi_i)\frac{Y_j - m_j}{\pi_j}\frac{Y_i - m_i}{\pi_i},$$

where

$$t_y^* = \sum_{j\in S}\frac{Y_j - m_j}{\pi_j} + \sum_{j\in U} m_j,$$

and m_j is the local polynomial regression estimator at point x_j, based on the entire finite population, given by

$$m_j = \mathbf{e}_1^T\left(\mathbf{X}_{Uj}^T\mathbf{W}_{Uj}\mathbf{X}_{Uj}\right)^{-1}\mathbf{X}_{Uj}^T\mathbf{W}_{Uj}\mathbf{Y}_U = \mathbf{w}_{Uj}^T\mathbf{Y}_U. \tag{3.3}$$

In expression (3.3), $\mathbf{X}_{Uj}$ and $\mathbf{W}_{Uj}$ are defined as follows:

$$\mathbf{X}_{Uj} = \begin{pmatrix} 1 & (x_1 - x_j) & \cdots & (x_1 - x_j)^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (x_N - x_j) & \cdots & (x_N - x_j)^q \end{pmatrix},$$

$$\mathbf{W}_{Uj} = \mathrm{diag}\left\{K\left(\frac{x_i - x_j}{h}\right)\frac{1}{h},\quad i \in U\right\}.$$

Breidt and Opsomer (2000) also showed that the MSE of $\hat{t}_y$ is consistently estimated by

$$\hat{\mathrm{V}}(\hat{t}_y) = \sum_{j\in S}\sum_{i\in S}\frac{\pi_{ji} - \pi_j\pi_i}{\pi_{ji}}\frac{Y_j - \hat{m}_j}{\pi_j}\frac{Y_i - \hat{m}_i}{\pi_i}. \tag{3.4}$$

3.2 Model averaging

In practical survey estimation problems, especially large-scale ones, usually multiple response variables are of interest and many auxiliary variables are available. In order to get a good regression estimator in terms of both efficiency and simplicity, a natural approach is to use model selection procedures. However, despite the nice theoretical properties, model selection often results in rather unstable estimators in applications. A small variation of the data may produce a very different model. Therefore, regression estimators based on model selection often have unnecessarily large variance. In addition, when there are multiple study variables, it seems almost impossible to select one model that fits all the study variables well. We will show this in the simulation section of this chapter.

An alternative to model selection is model averaging. Intuitively, if two models are very close with respect to a selection criterion, proper weighting of the models can be better than choosing only one of them (an exaggerated $0-1$ decision). In this way, we can eliminate the uncertainty of model selection procedures. Various work has been done in the area of model mixing, such as Breiman (1996), LeBlanc and Tibshirani (1996) and Yang (2001). However, there are few applications in survey estimation.

In this work, we propose a model averaging estimator that can be properly applied to survey estimation problems. We focus on the local polynomial regression estimator $\hat{t}_y$ defined in (3.1) because nonparametric regression is flexible for a wide range of models and will not suffer from misspecifying the true model as much as parametric regression. The estimator (3.1) depends on the value of bandwidth h, so its MSE can be considered as a function of h. Selecting proper candidates for model averaging in this case is equivalent to selecting proper values of bandwidth h. See Opsomer and Miller (2005), for example, on optimal bandwidth selection. Note that estimator (3.1) can also be written in the form of a weighted sum of Y_j's, i.e. $\hat{t}_y = \sum_{j \in S} w_j^* Y_j$, so the weights w_j^* also depends on the value of bandwidth h and each set of regression weights correspond to a different regression model procedure.

Suppose we have a finite collection of regression procedures to estimate the regression function m. Let the proposed procedures be δ_k, $k = 1, \cdots, K$. Our goal is to provide a method that can properly mix the K regression procedures. The resulting model averaging estimator should be flexible and perform well for multiple study variables under a wide range of regression models. In other words, this model averaging method should be overall a good choice for all study variables. In practical survey problems, several sets of regression weights are often available, with each set being obtained from a certain regression procedure. Suppose we have K sets of regression weights, then for each study variable Y_j, there are K possible regression estimators for t_y, denoted by $\hat{t}_{yk}$, $k = 1, \cdots, K$. This model averaging estimator should be appealing due to its flexibility. It should also significantly reduce the amount of work needed for estimation because it does not require a separate estimation procedure for each individual study variable. All that is needed is to use this method to average the K sets of regression weights and apply it to all study variables.

We consider regression procedure δ_k to be local polynomial regression with bandwidth h_k. The corresponding regression estimator $\hat{t}_{yk}$ is

$$\hat{t}_{yk} = \sum_{j \in S} \frac{Y_j - \hat{m}_{j,k}}{\pi_j} + \sum_{j \in U} \hat{m}_{j,k}, \tag{3.5}$$

where $\hat{m}_{j,k}$ is the regression predictor for model function m, using procedure δ_k, and

$$\begin{aligned} \hat{m}_{j,k} &= \mathbf{e}_1^T \left(\mathbf{X}_{Sj}^T \mathbf{W}_{Sj,k} \mathbf{X}_{Sj} + \frac{\nu}{N^{q+1}} \mathbf{I} \right)^{-1} \mathbf{X}_{Sj}^T \mathbf{W}_{Sj,k} \mathbf{Y}_S \qquad (3.6) \\ &= \mathbf{w}_{Sj,k}^T \mathbf{Y}_S. \end{aligned}$$

Our proposed model averaging (MA) estimator is of the simple linear form

$$\hat{t}_y^{MA} = \sum_{k=1}^{K} \alpha_k \hat{t}_{yk}, \qquad (3.7)$$

where $\hat{t}_{yk}$ is defined in (3.5), $\alpha_k \geq 0$ and $\sum_{k=1}^{K} \alpha_k = 1$. Here α_k is the 'weight' that is assigned to procedure δ_k. We are interested in finding the appropriate α_k's for the model averaging estimator $\hat{t}_y^{MA}$. Note that

$$\begin{aligned} \hat{t}_y^{MA} &= \sum_{k=1}^{K} \alpha_k \hat{t}_{yk} \\ &= \sum_{k=1}^{K} \alpha_k \left(\sum_{j \in S} \frac{Y_j - \hat{m}_{j,k}}{\pi_j} + \sum_{j \in U} \hat{m}_{j,k} \right) \\ &= \sum_{j \in S} \sum_{k=1}^{K} \frac{\alpha_k Y_j - \alpha_k \hat{m}_{j,k}}{\pi_j} + \sum_{j \in U} \sum_{k=1}^{K} \alpha_k \hat{m}_{j,k} \\ &= \sum_{j \in S} \frac{Y_j - \sum_{k=1}^{K} \alpha_k \hat{m}_{j,k}}{\pi_j} + \sum_{j \in U} \sum_{k=1}^{K} \alpha_k \hat{m}_{j,k} \\ &= \sum_{j \in S} \frac{Y_j - \hat{m}_j^{MA}}{\pi_j} + \sum_{j \in U} \hat{m}_j^{MA}, \end{aligned}$$

where

$$\hat{m}_j^{MA} = \sum_{k=1}^{K} \alpha_k \hat{m}_{j,k}. \qquad (3.8)$$

So choosing the proper α_k's for $\hat{t}_{yk}$ is equivalent to choosing proper α_k's for $\hat{m}_{j,k}$.

To proceed with model averaging, let us first consider selecting one best model, i.e. the optimal bandwidth h_{opt}. As stated in section 3.1, Breidt and Opsomer (2000) showed that the MSE of $\hat{t}_y$ is consistently estimated by $\hat{\mathrm{V}}(\hat{t}_y)$, where $\hat{\mathrm{V}}(\hat{t}_y)$ is defined in equation (3.4). It seems tempting to consider that h_{opt} should minimize $\hat{\mathrm{V}}(\hat{t}_y)$. However, this is not true. One can always choose arbitrarily small bandwidth h so that $\hat{m}_i$ is as close to Y_i as possible. Therefore, as a modification, Opsomer and Miller (2005) proposed a design-based Cross-validation (CV) criterion:

$$\hat{\mathrm{V}}_{CV}(h_k) = \sum_{j\in S}\sum_{i\in S} \frac{\pi_{ji} - \pi_j\pi_i}{\pi_{ji}} \frac{Y_j - \hat{m}_{j,k}^{(-)}}{\pi_j} \frac{Y_i - \hat{m}_{i,k}^{(-)}}{\pi_i}. \tag{3.9}$$

where $\hat{m}_{j,k}^{(-)}$ is the 'leave-one-out' CV estimator for m_j using procedure δ_k. To obtain this, we replace $\mathbf{w}_{Sj,k}$ in equation (3.6) by a modified vector $\mathbf{w}'_{Sj,k}$, whose elements are

$$w'_{Sji,k} = \begin{cases} \frac{w_{Sji,k}}{1-w_{Sji,k}} & \text{if} \quad j \neq i \\ 0 & \text{if} \quad j = i, \end{cases}$$

where $w_{Sji,k}$ denotes the jth element of the vector $\mathbf{w}_{Sj,k}$, and set $\hat{m}_{j,k}^{(-)} = \sum_{j\in S} w'_{Sji,k} Y_j$.

For model averaging purposes, we will consider the following methods to combine models:

1. Take the average of C estimators that have the lowest C $\hat{V}_{CV}(h_k)$.

2. Choose the estimator with the lowest $\hat{V}_{CV}(h_k)$, then we also include estimators with $\hat{V}_{CV}(h_k)$ that are within, say, a $p\%$ window above the lowest one. Then we take average of these estimators.

3. LeBlanc and Tibshirani (1996) suggested using

   $$\alpha_k = \frac{\hat{\sigma}_k^{-n}}{\sum_{j=1}^K \hat{\sigma}_j^{-n}}$$

   for a normal model, where $\hat{\sigma}_k^2$ is the resubstitution estimate of prediction error for model k. We consider a slightly different model averaging coefficient, α_k^*, where

   $$\alpha_k^* = \frac{\hat{\sigma}_k^{-1}}{\sum_{j=1}^K \hat{\sigma}_j^{-1}},$$

   and $\hat{\sigma}_k^2$ is defined as

   $$\hat{\sigma}_k^2 = \frac{1}{n}\sum_{j\in S}(Y_j - \hat{m}_{j,k}^{(-)})^2.$$

4. With constraint $\alpha_k \geq 0$ and $\sum_{k=1}^{K} \alpha_k = 1$, choose α_k to minimize the following criteria:

$$\hat{\sigma}^2 = \sum_{j \in S} \frac{1}{\pi_j} (Y_j - \hat{m}_j^{MA})^2, \tag{3.10}$$

$$\hat{\sigma}^2_{CV} = \sum_{j \in S} \frac{1}{\pi_j} (Y_j - \hat{m}_j^{MA(-)})^2, \tag{3.11}$$

$$\hat{\mathrm{V}}(t_y^{MA}) = \sum_{j \in S} \sum_{i \in S} \frac{\pi_{ji} - \pi_j \pi_i}{\pi_{ji}} \frac{Y_j - \hat{m}_j^{MA}}{\pi_j} \frac{Y_i - \hat{m}_i^{MA}}{\pi_i}, \tag{3.12}$$

$$\hat{\mathrm{V}}_{CV}(t_y^{MA}) = \sum_{j \in S} \sum_{i \in S} \frac{\pi_{ji} - \pi_j \pi_i}{\pi_{ji}} \frac{Y_j - \hat{m}_j^{MA(-)}}{\pi_j} \frac{Y_i - \hat{m}_i^{MA(-)}}{\pi_i}. \tag{3.13}$$

Equation (3.12) uses a similar idea of minimizing $\hat{\mathrm{V}}(\hat{t}_y)$ in (3.4). Equation (3.13) borrows the idea of the CV criterion in (3.9). Equation (3.10) and (3.11) are similar to (3.12) and (3.13), respectively, except that they do not fully incorporate the sampling design. Among equation (3.10) to (3.13), (3.13) will probably provide the best estimator for population total t, but it is the most computational intensive one. So we also investigate (3.10) to (3.12). Equation (3.10) requires the least amount of computation, if it works decently well, we may choose it over other methods. However, as we have discussed before, (3.10) can be minimized by choosing arbitrarily small bandwidth values. So it will probably not produce a good estimator. Same reasoning applies to (3.12). Equation (3.11) is an improved version of (3.10). But neither (3.10) or (3.11) fully incorporate the design.

In the following sections, we will illustrate the properties of different model averaging estimators through a large-scale simulation study.

3.3 Simulation setup

To evaluate the properties of Model Averaging (MA) estimators, we generate a single finite population and draw samples repeatedly from it. Specifically, we generate $N = 2000$ values of model variable X from the uniform distribution on $[0, 1]$, and 2000 values of error ε from $N(0, 1)$. This set of errors are used for all populations, up to multiplication by σ. We examine eight populations of Y:

$$Y_{jl} = m_l(x_j) + \varepsilon_j, \quad 1 \leq i \leq 2000, \quad 1 \leq l \leq 8$$

where $m_l(x_j)$ are defined on the third column of Table 3.1. We vary the value of σ to achieve high and low *coefficient of determination*, denoted by R^2. Specifically, we let $R^2 = 0.75$ and 0.25, where R^2 is defined as

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{j \in U} \varepsilon_j^2}{\sum_{j \in U} (Y_j - \bar{Y}_U)^2}.$$

We also consider two sampling designs. One is simple random sampling without replacement (SRS) and the other is random stratified sampling (STSRS), that is, we draw an SRS sample within each stratum. We choose two sample sizes, $n = 500$ and $n = 100$. It is easy to see that for SRS design, each element has equal selection probability, which is n/N. For STSRS design, we assign a different selection probability to each stratum. Specifically, we create two equally sized strata in each finite population by the values of model variable x. Then from the first stratum, we draw $n/4$ points and from the second stratum, we draw $3n/4$ points. So the ratio of selection probabilities are 1:3 for these two strata.

For model averaging estimation purpose, we examine five regression procedures. Specifically, we consider local polynomial regression with five different bandwidth values. We choose bandwidth from 0.01 to 0.5, equally spaced on the natural logarithm scale. These values are $h_1 = 0.01$, $h_2 = 0.027$, $h_3 = 0.071$, $h_4 = 0.188$, and $h_5 = 0.5$.

Name	Abbreviation	Expression
(1) Linear	LINE	$2x$
(2) Quadratic	QUAD	$1 + 2(x - 0.5)^2$
(3) Bump	BUMP	$2x + \exp(-200(x - 0.5)^2)$
(4) Jump	JUMP	$\begin{cases} 2x & \text{if} \quad x \le 0.65 \\ 0.65 & \text{if} \quad x > 0.65 \end{cases}$
(5) Normal CDF	NCDF	$\Phi^{-1}(1.5 - 2x)$
(6) Exponential	EXPO	$\exp(-8x)$
(7) Slow sine	SLOW	$2 + \sin(2\pi x)$
(8) Fast sine	FAST	$2 + \sin(8\pi x)$

Table 3.1: List of population functions

The finite population quantities of interest are $t_{yl} = \sum_{j \in U} Y_{jl}$ for each l. For each simulation, $B = 10000$ samples are drawn from each population. For each sample, we obtain five different estimators for t, denoted by $\{\hat{t}_{yk}\}_{k=1}^{5}$, where $\hat{t}_{yk}$ is defined in (3.5). Then we consider different methods to compute model averaging estimator $\hat{t}_y^{MA}$ described in the previous section. The details are listed in Table 3.2. Note that the last five rows in Table 3.2 are simply $\{\hat{t}_{yk}\}_{k=1}^{5}$. We list them here mainly for two reasons. One is to understand the behavior of each regression procedure, and the other is to compare them with other estimators to see if there are advantages to use model averaging. Loosely speaking, we will call all 13 estimators listed in Table 3.2 model averaging estimators. The last five rows can be regarded as model averaging of one regression procedure.

In summary, there are eight mean functions, two coefficients of determination ($R^2 =$ 0.75 and 0.25), two sampling designs (SRS and STSRS), two sample sizes ($n = 500$ and 100), and 13 estimators for each population total.

Relative Bias (RB) and Mean Squared Error (MSE) are computed for each estimator. Let $\{\hat{t}_{yr}^{MA}\}_{r=1}^{13}$ denote the thirteen estimators listed in Table 3.2, then

$$\begin{aligned} \mathrm{RB}_r &= \frac{\mathrm{E}(\hat{t}_{yr}^{MA}) - t_y}{t_y}, \\ \text{and} \quad \mathrm{MSE}_r &= \mathrm{E}(\hat{t}_{yr}^{MA} - t_y)^2. \end{aligned}$$

Method ID	Description
CV	Choose the estimator that has the lowest $\hat{V}_{CV}(h_k)$.
CV3	Take the average of 3 estimators that have the lowest 3 $\hat{V}_{CV}(h_k)$.
CVp20	Choose the estimator that has the lowest $\hat{V}_{CV}(h_k)$. Also include estimators having $\hat{V}_{CV}(h_k)$ that are $\leq 20\%$ bigger than the lowest one.
Relative Fit	Use α_k's as described in method 3. i.e. $\alpha_k = \frac{\hat{\sigma}_k^{-1}}{\sum_{j=1}^{K}\hat{\sigma}_j^{-1}}$
MIN$\{\hat{\sigma}^2\}$	With constraint $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$, choose α_k to minimize (3.10).
MIN$\{\hat{\sigma}^2_{CV}\}$	With constraint $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$, choose α_k to minimize (3.11).
MIN$\{\hat{\mathrm{V}}\}$	With constraint $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$, choose α_k to minimize (3.12).
MIN$\{\hat{\mathrm{V}}_{CV}\}$	With constraint $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$, choose α_k to minimize (3.13).
FIXt(0.010)	Choose the estimator that uses bandwidth $h_1 = 0.010$.
FIXt(0.027)	Choose the estimator that uses bandwidth $h_2 = 0.027$.
FIXt(0.071)	Choose the estimator that uses bandwidth $h_3 = 0.071$.
FIXt(0.188)	Choose the estimator that uses bandwidth $h_4 = 0.188$.
FIXt(0.500)	Choose the estimator that uses bandwidth $h_5 = 0.500$.

Table 3.2: List of case IDs and corresponding descriptions.

3.4 Simulation results

Table 3.3 to Table 3.6 report the simulated relative bias (in percent) of 13 estimators for eight population totals where SRS samples are drawn to compute $\hat{t}_{yr}^{MA}$. These four tables correspond to the following four scenarios: (1) $R^2 = 0.75$, $n = 500$; (2) $R^2 = 0.25$, $n = 500$; (3) $R^2 = 0.75$, $n = 100$; (4) $R^2 = 0.25$, $n = 100$. We can see that all the relative biases in these tables are very small. When $R^2 = 0.75$ and $n = 500$, almost all $\hat{t}_{yr}^{MA}$ are essentially unbiased. As the sample size n gets smaller and the coefficient of determination R^2 gets smaller, we see an increasing trend of relative bias from Table 3.3 to Table 3.6. In Table 3.5 and Table 3.6, except for population 'Normal CDF', all other populations have negative biases for most of the 13 model averaging total estimators. Also, the relative bias for population 'Exponential' in Table 3.5 and Table 3.6 is substantially higher than other populations. This fact suggests that when the population function is exponential, we need larger sample size to achieve similar amount of relative bias to other populations. We can also see that in most cases, MIN$\{\hat{\sigma}^2_{CV}\}$

and MIN$\{\hat{\mathrm{V}}_{CV}\}$ have similar relative biases, and they are lower than those of MIN$\{\hat{\sigma}^2\}$ and MIN$\{\hat{\mathrm{V}}\}$. This indicates that in order to calculate α_k for $\hat{t}_y^{MA}$, using 'leave-one-out' estimator $\hat{m}_j^{(-)}$ is slightly better than using $\hat{m}_j$ as far as relative bias is concerned. By looking at the last five rows of each table, we can see that when $n = 500$, the smallest bias usually occurs when h is near 0.071. When sample size decreases to 100, local polynomial regression tends to favor slightly bigger bandwidth, that is, around 0.188. Note that neither the smallest bandwidth 0.01 nor the biggest one 0.5 is the best choice with respect to relative bias. If we only compare the first eight model averaging methods (the last five rows in Table 3.3 to Table 3.6 are not exactly model averaging), we do not see much difference among different estimators in terms of relative bias, although we expect MIN$\{\hat{\mathrm{V}}_{CV}\}$ to be the best model averaging method.

Table 3.7 to Table 3.10 report the simulated relative bias (in percent) of 13 estimators for eight population totals where STSRS samples are drawn to compute $\hat{t}_{yr}^{MA}$. As we can see, most relative biases in Table 3.7 to Table 3.9 are larger than those in Table 3.3 to Table 3.5, correspondingly. In Table 3.10, population 'Quadratic', 'Exponential' and 'Fast sine' have larger biases than the corresponding ones in Table 3.6. For the other five populations, the relative biases are slightly smaller than those in Table 3.6. In Table 3.9 and Table 3.10, the relative biases for population 'Exponential' are substantially larger than those for other populations (more than 1% vs less than 0.1%). We expect the averaging method MIN$\{\hat{\mathrm{V}}_{CV}\}$ to be the best, however, if we only compare the first eight rows in Table 3.7 to Table 3.10, we can not see a big difference among different averaging methods in terms of relative bias.

Based on our observations from Table 3.3 to Table 3.10, as far as relative bias is concerned, all model averaging methods perform similarly well. For most cases, method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ produce less biases than MIN$\{\hat{\sigma}^2\}$ and MIN$\{\hat{\mathrm{V}}\}$.

In order to evaluate the performance of different model averaging estimators, we also examine their variabilities by looking at their mean squared errors (MSE).

Table 3.11 to Table 3.14 report the MSEs of 12 estimators relative to the MSE of method CV, and minus one (in percent) for eight population totals where SRS samples are drawn to calculate $\hat{t}_{yr}^{MA}$. Specifically, the values in Table 3.11 to Table 3.14 are calculated as follows:

$$\frac{\text{MSE(12 estimators)}}{\text{MSE(CV)}} - 1.$$

Note that this quantity shows how much higher (positive values) or lower (negative values) of one method relative to method CV in terms of MSE. For example, if a value is 50, it means that the correspond method's MSE is 50% higher than that of method CV. These four tables correspond to the following four senarios: (1) $R^2 = 0.75$, $n = 500$; (2) $R^2 = 0.25$, $n = 500$; (3) $R^2 = 0.75$, $n = 100$; (4) $R^2 = 0.25$, $n = 100$. The first thing we can notice is that the differences among different model averaging estimators for each population total are bigger than those of relative biases. If we examine these tables more closely, we can observe the following facts:

1. In all cases, method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ have smaller MSEs than method MIN$\{\hat{\sigma}^2\}$ and MIN$\{\hat{\mathrm{V}}\}$. So it is better to use leave-one-out CV estimator for m_j in terms of the variability of model averaging estimation.

2. In most cases, MIN$\{\hat{\sigma}^2_{CV}\}$ is slightly better than, or at least as good as MIN$\{\hat{\mathrm{V}}_{CV}\}$ in terms of MSE. Few exceptions exist where MIN$\{\hat{\mathrm{V}}_{CV}\}$ is better than MIN$\{\hat{\sigma}^2_{CV}\}$ with respect to MSE. They are: population 'Quadratic' and 'Normal CDF' in Table 3.13 and population 'Linear', 'Normal CDF' and 'Exponential' in Table 3.14.

3. Method CV, CV3, CVp20, MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ return similar MSEs. In Table 3.11 and Table 3.12, where sample size is 500, the MSEs of these five model averaging estimators are closer to each other than those in Table 3.13 and Table 3.14, where the sample size is 100. So when the sample size is large enough, it becomes very hard to choose among these five model averaging estimator. Now

let us focus on smaller sample size. In Table 3.13 and Table 3.14, if we compare method CV, CV3 and CVp20, we cannot choose the universally best among these three methods. For example, in Table 3.13, CV has the smallest MSE among CV, CV3 and CVp20, for population 'Linear', 'Quadratic', 'Bump' and 'Exponential' (positive values), but for population 'Jump', 'Normal CDF' and 'Fast sine', CV has the largest MSE among CV, CV3 and CVp20 (negative values). If we also consider MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$, we can see that these two methods give very similar MSEs, and are either the smallest among CV, CV3, CVp20, MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ or close to the smallest one. We like the fact that method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ are consistently good. Other methods can be inconsistent. They can either be the best for a certain population, or the worst for another. For instance, in Table 3.13, among method CV, CV3, CVp20, MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$, CV is the best for population 'Bump', but the worst for population 'Fast sine'.

4. The method Relative Fit behaves well in Table 3.11 and Table 3.12, where the sample size is 500. There is one exception, however. In Table 3.11, for population 'Fast sine', the MSE of method Relative Fit is 75.37% higher than the MSE of method CV. When the sample size decreases to 100, i.e. in Table 3.13 and Table 3.14, it starts to show larger and larger MSEs than method CV, CV3, CVp20, MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$, but still smaller than method MIN$\{\hat{\sigma}^2\}$ and MIN$\{\hat{\mathrm{V}}\}$.

5. From the last five rows in Table 3.11 to Table 3.14, we can see that the MSEs of the five regression procedures vary greatly for each population , especially when sample size is small ($n = 100$). Specifically, a regression procedure can be very good for a certain population, but very bad for another. For example, in Table 3.13, FIXt(0.188) has the smallest MSE for population 'Normal CDF' (3.5% lower than CV) and 'Exponential' (1.7% lower than CV), but for population 'Fast sine',

it is almost the worst (151.2% higher than CV), where CVp20 (4.35% lower than CV) has the smallest MSE among all 13 estimators.

Table 3.15 to Table 3.18 report the MSEs of 13 estimators for eight population totals where STSRS samples are drawn to calculate $\hat{t}_{yr}^{MA}$. We can see that all the MSEs are larger than the corresponding ones in Table 3.11 to Table 3.14. We can also see that $\mathrm{MIN}\{\hat{\sigma}^2_{CV}\}$ has smaller or equal MSEs to $\mathrm{MIN}\{\hat{\mathrm{V}}_{CV}\}$ in more cases than SRS design, with only one exception here, that is, population 'Normal CDF' in Table 3.17. In the cases where $\mathrm{MIN}\{\hat{\sigma}^2_{CV}\}$ is better than $\mathrm{MIN}\{\hat{\mathrm{V}}_{CV}\}$, the differences are getting larger than those in Table 3.11 to Table 3.14.

Now we have shown the good properties of method $\mathrm{MIN}\{\hat{\sigma}^2_{CV}\}$ and $\mathrm{MIN}\{\hat{\mathrm{V}}_{CV}\}$. To further study their behaviors, we examine on a detailed level that how different model averaging methods combine regression procedures, i.e. we examine the mean values of α_k's and their standard errors over the $B = 10000$ replications. Now only the first eight methods are of interest because the fixed bandwidth ones are not indeed model averaging. Among the first eight methods, CV, Relative Fit, $\mathrm{MIN}\{\hat{\sigma}^2_{CV}\}$, and $\mathrm{MIN}\{\hat{\mathrm{V}}_{CV}\}$ are of particular interest. Method CV selects only one model for each replication, by comparing its α_k's with α_k's of method $\mathrm{MIN}\{\hat{\sigma}^2_{CV}\}$ and $\mathrm{MIN}\{\hat{\mathrm{V}}_{CV}\}$, we can have a better understanding whether method $\mathrm{MIN}\{\hat{\sigma}^2_{CV}\}$ and $\mathrm{MIN}\{\hat{\mathrm{V}}_{CV}\}$ choose one model a time or actually take the weighted average of several models. Relative Fit uses a completely different approach to obtain α_k's, so it is interesting to see what its α_k's are. Method $\mathrm{MIN}\{\hat{\sigma}^2\}$ and $\mathrm{MIN}\{\hat{\mathrm{V}}\}$ always assign $\alpha_1 = 1$ and other α_k's zero, i.e. they always select the model with smallest bandwidth. The results will not be listed here. But this fact explains why method $\mathrm{MIN}\{\hat{\sigma}^2\}$ and $\mathrm{MIN}\{\hat{\mathrm{V}}\}$ do not perform well in terms of MSE. If the bandwidth is too small, the nonparametric regression will produce a curve that is very 'wriggly' because it tries to capture the random errors into the whole trend. As a result, the shape of the smoothed curves will largely depend on the random errors.

Therefore the estimated population total will be highly variable.

Table 3.19 to Table 3.26 report the mean of model averaging coefficients $\{\alpha_k\}_{k=1}^5$ and corresponding standard errors over the B=10000 replications, using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for four combinations of sample size n and coefficient of determination R^2 and eight populations, with the sample design being SRS in these eight tables. Note that α_k corresponds to the weight assigned to local polynomial regression with bandwidth h_k, where $h_1 = 0.01$ is the smallest bandwidth and $h_5 = 0.5$ is the largest one.

For method CV, note that its α_k's are the proportions that bandwidth h_k are selected over the 10000 replications. For example, in Table 3.19, when $n = 500$ and $R^2 = 0.75$, $\alpha_5 = 1.00$ with standard error 0.02 for method CV. This means that *almost* 100% of the time $h_5 = 0.5$ is selected for linear population of $R^2 = 0.75$ and sample size = 500. As we can see, method CV tends to choose one particular model almost all the time for each combination of n and R^2, i.e. a very large α_k for a certain model and very small ones (mostly zeros) for others. There are two exceptions, though. In Table 3.22 and Table 3.25, when $n = 100$ and $R^2 = 0.75$, method CV almost evenly chooses between h_4 and h_5 (α_4 and α_5 are close to 0.5).

For method Relative Fit, we can see that α_1 to α_5 are all close to 0.2 with very small standard errors. This means that method Relative Fit is similar to just taking the average of all models.

As for method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$, we can see that they have similar α_k's most of the time, and their α_k's are quite different than method CV. Because method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ use basically the same criterion to choose models as CV, so if they choose only one model or mostly one model in each replication, the results should be similar to those for CV. The difference between α_k's for CV and α_k's for method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ suggests that method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ use more than one model to construct the model averaging estimator. Now we see that the good properties of method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ are due to the fact that they use more than one regression procedure to eliminate the uncertainty of model selection procedures.

Now let us examine more closely how method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ assign model averaging coefficients (α_k's) to different procedures. For population 'Linear' (Table 3.19), we can see that the procedure that has the largest model averaging coefficient is the one with $h = 0.5$, for all four combinations of n and R^2, where α_5 is around 0.75. For the other four regression procedures, the model averaging coefficients tend to increase as the corresponding bandwidth h increases. For all other seven populations, which correspond to Table 3.20 to Table 3.26, we can see that method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ tend to assign most of the model averaging coefficients to the mid-sized bandwidth (h_2, h_3 and h_4). Also, as n and R^2 become smaller, method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ tend to choose bigger bandwidths, i.e. as we look from the top to the bottom in Table 3.20 to Table 3.26, bigger α values tend to shift towards the right side of those tables.

For the unequal selection probability design STSRS, the α_k's for method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ are shown in Table 3.27 to Table 3.34. As for method CV, it selects a particular model less often than the SRS design, i.e. more α_k's have nonzero values larger standard errors. The α_k's for method Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ are very similar to those in Table 3.19 to Table 3.26, which use SRS, the equal selection probability design.

3.5 Simulation conclusions

From the previous statements, we can draw the following conclusions from this simulation study.

1. As far as biases are concerned, all 13 estimators perform very well. It is hard to

choose from them if we only consider their biases.

2. In terms of MSEs, if sample size is large ($n = 500$), model selection (CV) is important because it performs as well as the best available model averaging methods. If sample size is smaller ($n = 100$), we consider method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ to be the overall best choices, with MIN$\{\hat{\sigma}^2_{CV}\}$ being even better. Those two methods behave well in all cases. Method CV, CV3 and CVp20 are good, but they are not as consistent as method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for different cases and different study variables.

3. We can draw the same conclusions for both equal selection probability design, SRS, and unequal selection probability design, STSRS. The only difference we can see is that relative biases and MSEs for SRS are smaller than the corresponding ones for STSRS.

4. If model averaging is carried out in a proper way, it can eliminate uncertainty of model selection procedures and produce more reliable estimators. If one is to draw a sample to estimate the population total but do not have prior knowledge about the population model, we recommend the method MIN$\{\hat{\sigma}^2_{CV}\}$ and MIN$\{\hat{\mathrm{V}}_{CV}\}$ to be the safer bets than other methods because they give good estimators under all cases.

Averaging Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
	Population Functions							
CV	0.00	0.00	-0.01	-0.01	0.00	-0.01	0.00	0.00
CV3	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00
CVp20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Relative Fit	0.00	0.00	-0.01	-0.01	0.00	-0.01	0.00	0.01
MIN$\{\hat{\sigma}^2\}$	0.01	0.00	0.01	0.01	-0.02	0.04	0.01	0.01
MIN$\{\hat{\sigma}^2_{CV}\}$	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00
MIN$\{\hat{\mathrm{V}}\}$	0.01	0.00	0.01	0.01	-0.02	0.04	0.01	0.01
MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00
FIXt(0.010)	0.01	0.00	0.01	0.01	-0.02	0.04	0.01	0.01
FIXt(0.027)	-0.01	0.00	-0.01	0.00	0.01	-0.02	0.00	0.00
FIXt(0.071)	0.00	0.00	0.00	-0.01	0.00	-0.01	0.00	0.00
FIXt(0.188)	0.00	0.00	-0.01	-0.01	0.00	-0.01	0.00	0.01
FIXt(0.500)	0.00	0.00	-0.03	-0.04	-0.01	-0.08	0.01	0.01

Table 3.3: Relative Bias (in percent) for eight populations ($R^2 = 0.75$), Simple Random Sampling (SRS) with sample size $n = 500$ and thirteen model averaging methods.

Averaging Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
	Population Functions							
CV	-0.01	0.00	-0.02	-0.02	0.04	-0.04	0.00	0.00
CV3	-0.01	0.00	-0.04	-0.02	0.04	-0.07	-0.01	0.00
CVp20	0.00	0.00	-0.01	-0.01	0.04	-0.02	0.00	0.00
Relative Fit	0.00	0.00	-0.01	-0.01	0.04	-0.01	0.00	0.01
MIN$\{\hat{\sigma}^2\}$	0.04	0.01	0.04	0.03	0.02	0.13	0.02	0.02
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.01	0.00	-0.02	-0.01	0.04	-0.04	0.00	0.00
MIN$\{\hat{\mathrm{V}}\}$	0.04	0.01	0.04	0.03	0.02	0.13	0.02	0.02
MIN$\{\hat{\mathrm{V}}_{CV}\}$	-0.01	0.00	-0.02	-0.01	0.04	-0.05	0.00	0.00
FIXt(0.010)	0.04	0.01	0.04	0.03	0.02	0.13	0.02	0.02
FIXt(0.027)	-0.02	0.00	-0.02	-0.01	0.05	-0.05	-0.01	-0.01
FIXt(0.071)	-0.01	0.00	-0.01	-0.01	0.04	-0.03	-0.01	0.00
FIXt(0.188)	-0.01	0.00	-0.02	-0.02	0.04	-0.03	0.00	0.01
FIXt(0.500)	0.00	0.00	-0.04	-0.04	0.04	-0.09	0.01	0.01

Table 3.4: Relative Bias (in percent) for eight populations ($R^2 = 0.25$), Simple Random Sampling (SRS) with sample size $n = 500$ and thirteen model averaging methods.

Averaging Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
	Population Functions							
CV	-0.02	-0.01	-0.05	-0.02	0.01	-0.29	-0.01	-0.02
CV3	-0.02	-0.02	-0.11	-0.03	0.00	-0.50	-0.02	-0.01
CVp20	-0.02	-0.02	-0.08	-0.02	0.00	-0.48	-0.01	-0.02
Relative Fit	-0.03	-0.01	-0.02	-0.04	0.00	-0.33	-0.01	-0.02
MIN$\{\hat{\sigma}^2\}$	-0.02	-0.02	0.00	-0.02	-0.03	-0.36	-0.01	0.00
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.02	-0.01	-0.07	-0.03	0.00	-0.30	-0.01	-0.02
MIN$\{\hat{\text{V}}\}$	-0.02	-0.02	0.00	-0.02	-0.03	-0.36	-0.01	0.00
MIN$\{\hat{\text{V}}_{CV}\}$	-0.02	-0.01	-0.07	-0.03	0.00	-0.31	-0.01	-0.02
FIXt(0.010)	-0.02	-0.02	0.00	-0.02	-0.03	-0.36	-0.01	0.00
FIXt(0.027)	-0.06	-0.03	-0.05	-0.03	0.05	-0.52	-0.03	-0.02
FIXt(0.071)	-0.03	-0.02	-0.02	0.00	0.01	-0.34	-0.02	-0.02
FIXt(0.188)	-0.02	0.00	0.01	-0.02	0.00	-0.22	0.00	-0.03
FIXt(0.500)	-0.01	0.01	-0.06	-0.12	-0.02	-0.26	0.02	-0.02

Table 3.5: Relative Bias (in percent) for eight populations ($R^2 = 0.75$), Simple Random Sampling (SRS) with sample size $n = 100$ and thirteen model averaging methods.

Averaging Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
	Population Functions							
CV	-0.05	-0.03	-0.12	-0.08	0.11	-0.57	-0.03	-0.06
CV3	-0.06	-0.02	-0.07	-0.08	0.11	-0.42	-0.03	-0.08
CVp20	-0.07	-0.02	-0.08	-0.09	0.12	-0.46	-0.03	-0.07
Relative Fit	-0.08	-0.03	-0.08	-0.08	0.14	-0.53	-0.04	-0.05
MIN$\{\hat{\sigma}^2\}$	-0.05	-0.03	-0.04	-0.05	0.15	-0.48	-0.03	-0.02
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.06	-0.02	-0.12	-0.09	0.11	-0.54	-0.03	-0.06
MIN$\{\hat{\text{V}}\}$	-0.05	-0.03	-0.04	-0.05	0.15	-0.48	-0.03	-0.02
MIN$\{\hat{\text{V}}_{CV}\}$	-0.06	-0.02	-0.12	-0.08	0.11	-0.54	-0.03	-0.06
FIXt(0.010)	-0.05	-0.03	-0.04	-0.05	0.15	-0.48	-0.03	-0.02
FIXt(0.027)	-0.18	-0.06	-0.17	-0.12	0.21	-0.92	-0.10	-0.10
FIXt(0.071)	-0.09	-0.03	-0.08	-0.05	0.13	-0.54	-0.05	-0.05
FIXt(0.188)	-0.06	-0.01	-0.04	-0.06	0.11	-0.36	-0.03	-0.06
FIXt(0.500)	-0.03	0.00	-0.08	-0.14	0.09	-0.34	0.01	-0.03

Table 3.6: Relative Bias (in percent) for eight populations ($R^2 = 0.25$), Simple Random Sampling (SRS) with sample size $n = 100$ and thirteen model averaging methods.

Averaging				Population Functions				
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV	0.03	0.01	0.03	0.03	-0.01	0.04	0.01	0.03
CV3	0.03	0.01	0.03	0.03	-0.01	0.05	0.01	0.04
CVp20	0.03	0.01	0.03	0.03	-0.01	0.05	0.01	0.03
Relative Fit	0.03	0.01	0.03	0.03	-0.01	0.05	0.01	0.02
MIN$\{\hat{\sigma}^2\}$	0.04	0.01	0.04	0.03	-0.02	0.08	0.02	0.04
MIN$\{\hat{\sigma}^2_{CV}\}$	0.03	0.01	0.03	0.03	-0.01	0.04	0.01	0.03
MIN$\{\hat{\text{V}}\}$	0.04	0.01	0.04	0.03	-0.02	0.08	0.02	0.04
MIN$\{\hat{\text{V}}_{CV}\}$	0.03	0.01	0.03	0.03	-0.01	0.05	0.01	0.03
FIXt(0.010)	0.04	0.01	0.04	0.03	-0.02	0.08	0.02	0.04
FIXt(0.027)	0.03	0.01	0.03	0.03	-0.01	0.05	0.02	0.03
FIXt(0.071)	0.03	0.00	0.03	0.03	-0.01	0.04	0.01	0.04
FIXt(0.188)	0.03	0.01	0.04	0.03	-0.01	0.05	0.00	0.03
FIXt(0.500)	0.03	0.01	0.01	0.01	-0.02	0.02	0.01	-0.01

Table 3.7: Relative Bias (in percent) for eight populations ($R^2 = 0.75$), Random Stratified Sampling (STSRS) with sample size $n = 500$ and thirteen model averaging methods.

Averaging				Population Functions				
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV	0.08	0.02	0.06	0.07	-0.04	0.20	0.03	0.07
CV3	0.08	0.02	0.06	0.06	-0.04	0.18	0.04	0.07
CVp20	0.08	0.02	0.08	0.07	-0.04	0.21	0.04	0.06
Relative Fit	0.09	0.02	0.09	0.07	-0.04	0.23	0.04	0.06
MIN$\{\hat{\sigma}^2\}$	0.12	0.02	0.11	0.09	-0.05	0.32	0.07	0.08
MIN$\{\hat{\sigma}^2_{CV}\}$	0.09	0.02	0.07	0.07	-0.05	0.19	0.04	0.07
MIN$\{\hat{\text{V}}\}$	0.12	0.02	0.11	0.09	-0.05	0.32	0.07	0.08
MIN$\{\hat{\text{V}}_{CV}\}$	0.09	0.02	0.06	0.07	-0.05	0.20	0.04	0.07
FIXt(0.010)	0.12	0.02	0.11	0.09	-0.05	0.32	0.07	0.08
FIXt(0.027)	0.09	0.02	0.09	0.08	-0.04	0.23	0.05	0.07
FIXt(0.071)	0.08	0.02	0.08	0.07	-0.03	0.20	0.04	0.07
FIXt(0.188)	0.08	0.02	0.08	0.07	-0.04	0.20	0.03	0.05
FIXt(0.500)	0.09	0.02	0.07	0.05	-0.04	0.19	0.03	0.02

Table 3.8: Relative Bias (in percent) for eight populations ($R^2 = 0.25$), Random Stratified Sampling (STSRS) with sample size $n = 500$ and thirteen model averaging methods.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV	0.06	-0.04	0.00	0.08	0.02	-1.20	0.00	0.21
CV3	0.05	-0.04	-0.03	0.07	0.00	-1.42	0.02	0.25
CVp20	0.06	-0.05	-0.02	0.07	0.00	-1.34	0.01	0.23
Relative Fit	0.04	-0.04	0.08	0.06	-0.02	-1.18	-0.01	0.11
MIN$\{\hat{\sigma}^2\}$	0.05	-0.05	0.13	0.08	-0.10	-1.17	-0.01	0.30
MIN$\{\hat{\sigma}^2_{CV}\}$	0.07	-0.04	0.04	0.08	0.00	-1.20	-0.01	0.23
MIN$\{\hat{\mathrm{V}}\}$	0.05	-0.05	0.13	0.08	-0.10	-1.17	-0.01	0.30
MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	-0.04	-0.01	0.08	0.01	-1.18	0.00	0.21
FIXt(0.010)	0.05	-0.05	0.13	0.08	-0.10	-1.17	-0.01	0.30
FIXt(0.027)	0.02	-0.06	0.10	0.08	-0.02	-1.37	-0.02	0.27
FIXt(0.071)	0.02	-0.05	0.08	0.07	0.03	-1.37	0.00	0.19
FIXt(0.188)	0.05	-0.04	0.09	0.11	0.02	-1.06	-0.01	0.04
FIXt(0.500)	0.08	-0.03	-0.02	-0.03	-0.02	-1.02	-0.03	-0.08

Table 3.9: Relative Bias (in percent) for eight populations ($R^2 = 0.75$), Random Stratified Sampling (STSRS) with sample size $n = 100$ and thirteen model averaging methods.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV	0.01	-0.05	-0.07	0.01	0.01	-1.45	0.01	0.05
CV3	-0.02	-0.06	-0.02	0.00	0.01	-1.43	-0.05	0.05
CVp20	-0.02	-0.05	-0.03	0.00	0.05	-1.39	-0.05	0.06
Relative Fit	-0.04	-0.06	0.00	0.00	0.05	-1.50	-0.06	0.10
MIN$\{\hat{\sigma}^2\}$	-0.01	-0.06	0.07	0.04	0.02	-1.39	-0.05	0.27
MIN$\{\hat{\sigma}^2_{CV}\}$	0.02	-0.05	-0.03	0.02	0.01	-1.43	-0.04	0.09
MIN$\{\hat{\mathrm{V}}\}$	-0.01	-0.06	0.07	0.04	0.02	-1.39	-0.05	0.27
MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.01	-0.05	-0.06	-0.01	0.02	-1.39	-0.02	0.05
FIXt(0.010)	-0.01	-0.06	0.07	0.04	0.02	-1.39	-0.05	0.27
FIXt(0.027)	-0.10	-0.08	-0.02	-0.02	0.11	-1.80	-0.09	0.19
FIXt(0.071)	-0.10	-0.08	-0.04	-0.02	0.09	-1.77	-0.06	0.12
FIXt(0.188)	-0.01	-0.05	0.03	0.06	0.04	-1.29	-0.05	0.00
FIXt(0.500)	0.03	-0.04	-0.07	-0.07	0.00	-1.21	-0.06	-0.11

Table 3.10: Relative Bias (in percent) for eight populations ($R^2 = 0.25$), Random Stratified Sampling (STSRS) with sample size $n = 100$ and thirteen model averaging methods.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	-0.17	-0.28	0.70	-0.84	0.20	0.16	0.43	1.97
CVp20	0.76	0.64	1.00	-0.18	1.15	0.94	1.13	2.18
Relative Fit	0.90	2.24	4.11	4.70	1.34	2.39	4.90	75.37
MIN$\{\hat{\sigma}^2\}$	13.74	12.41	10.64	6.68	16.34	12.28	12.74	9.17
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.09	-0.33	-0.42	-0.55	0.23	0.08	0.18	0.12
MIN$\{\hat{\text{V}}\}$	13.74	12.41	10.64	6.68	16.34	12.28	12.74	9.17
MIN$\{\hat{\text{V}}_{CV}\}$	-0.10	-0.31	-0.42	-0.55	0.30	0.13	0.18	0.12
FIXt(0.010)	13.74	12.41	10.64	6.68	16.34	12.28	12.74	9.17
FIXt(0.027)	3.75	2.55	0.94	-0.27	3.73	2.44	2.84	0.01
FIXt(0.071)	0.75	-0.34	-0.45	3.00	0.60	-0.17	-0.08	16.26
FIXt(0.188)	-0.04	-0.17	16.95	14.40	-0.38	2.50	3.60	222.08
FIXt(0.500)	-0.44	31.35	37.07	55.10	2.10	30.59	58.52	230.60

Table 3.11: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.75$), Simple Random Sampling (SRS) with sample size $n = 500$.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	-0.10	0.27	0.23	-0.15	-0.26	0.01	-0.07	0.22
CVp20	0.84	0.62	0.26	-0.01	0.83	0.38	0.34	1.48
Relative Fit	0.98	0.73	0.31	0.03	0.97	0.48	0.41	2.72
MIN$\{\hat{\sigma}^2\}$	13.83	13.18	12.33	12.28	14.90	12.87	12.73	10.17
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.08	-0.14	0.05	-0.39	-0.03	-0.23	-0.33	-0.79
MIN$\{\hat{\text{V}}\}$	13.83	13.18	12.33	12.28	14.90	12.87	12.73	10.17
MIN$\{\hat{\text{V}}_{CV}\}$	-0.08	-0.11	0.05	-0.39	-0.04	-0.19	-0.33	-0.79
FIXt(0.010)	13.83	13.18	12.33	12.28	14.90	12.87	12.73	10.17
FIXt(0.027)	3.83	3.24	2.46	2.60	3.04	2.96	2.83	0.51
FIXt(0.071)	0.84	0.27	-0.27	0.08	0.43	0.05	-0.15	-0.55
FIXt(0.188)	0.04	-0.34	1.53	-0.53	-0.05	-0.35	-0.44	19.48
FIXt(0.500)	-0.42	3.09	3.12	3.65	-0.32	2.40	5.27	19.67

Table 3.12: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.25$), Simple Random Sampling (SRS) with sample size $n = 500$.

Averaging Method	Population Functions LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	0.17	3.57	2.66	-2.73	-2.40	1.19	0.46	-3.31
CVp20	0.80	0.95	2.29	-2.84	-1.91	-0.02	-1.61	-4.35
Relative Fit	7.61	5.96	2.42	-1.33	3.72	3.00	4.46	39.13
MIN$\{\hat{\sigma}^2\}$	38.77	34.71	25.99	18.92	33.15	29.96	29.21	9.38
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.14	0.09	0.45	-2.32	-0.96	-0.76	-1.36	-4.12
MIN$\{\hat{\mathrm{V}}\}$	38.77	34.71	25.99	18.92	33.15	29.96	29.21	9.38
MIN$\{\hat{\mathrm{V}}_{CV}\}$	-0.16	-0.11	0.49	-2.26	-1.04	-0.78	-1.34	-4.04
FIXt(0.010)	38.77	34.71	25.99	18.92	33.15	29.96	29.21	9.38
FIXt(0.027)	26.24	22.67	14.68	9.45	21.84	18.62	17.56	-0.15
FIXt(0.071)	6.76	3.79	-1.58	-3.07	3.05	0.87	-0.42	-0.15
FIXt(0.188)	0.37	-1.45	9.71	1.78	-3.50	-1.70	-1.50	151.20
FIXt(0.500)	-1.45	29.41	27.44	38.84	-3.02	23.27	50.58	159.56

Table 3.13: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.75$), Simple Random Sampling (SRS) with sample size $n = 100$.

Averaging Method	Population Functions LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	0.51	-2.29	-2.56	-1.47	-0.42	-1.77	-1.08	0.97
CVp20	1.12	-1.77	-2.22	-1.05	0.16	-1.16	-0.76	0.93
Relative Fit	8.02	4.55	2.96	4.43	6.61	5.09	4.92	0.88
MIN$\{\hat{\sigma}^2\}$	39.33	34.79	32.07	34.23	37.09	35.34	34.72	25.54
MIN$\{\hat{\sigma}^2_{CV}\}$	0.02	-1.12	-1.50	-0.35	-0.52	-0.56	-0.08	-0.59
MIN$\{\hat{\mathrm{V}}\}$	39.33	34.79	32.07	34.23	37.09	35.34	34.72	25.54
MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.01	-1.12	-1.49	-0.20	-0.53	-0.64	-0.07	-0.58
FIXt(0.010)	39.33	34.79	32.07	34.23	37.09	35.34	34.72	25.54
FIXt(0.027)	26.78	22.67	20.18	22.35	24.57	23.23	22.58	14.20
FIXt(0.071)	7.24	3.76	1.88	3.82	5.87	4.34	3.71	-1.63
FIXt(0.188)	0.72	-2.53	-1.73	-2.29	-0.17	-1.80	-1.92	10.27
FIXt(0.500)	-1.38	-1.36	-1.70	1.06	-2.04	-1.33	2.49	9.43

Table 3.14: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.25$), Simple Random Sampling (SRS) with sample size $n = 100$.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	-0.25	0.01	1.31	-1.78	0.29	-0.46	-0.56	1.66
CVp20	0.63	0.48	1.69	-1.08	1.20	0.09	-0.12	2.25
Relative Fit	2.23	2.68	4.16	-0.36	1.43	2.87	3.51	72.04
MIN$\{\hat{\sigma}^2\}$	23.38	21.56	18.61	15.36	12.45	19.85	20.70	13.78
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.02	-0.17	-0.34	0.17	0.15	-0.33	-0.50	0.01
MIN$\{\hat{\mathrm{V}}\}$	23.38	21.56	18.61	15.36	12.45	19.85	20.70	13.78
MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.09	-0.21	-0.11	-0.97	0.15	-0.10	-0.45	0.13
FIXt(0.010)	23.38	21.56	18.61	15.36	12.45	19.85	20.70	13.78
FIXt(0.027)	7.90	6.34	3.75	2.10	3.33	4.91	5.52	-0.15
FIXt(0.071)	1.31	-0.15	-0.94	-1.39	0.46	-0.99	-0.84	14.87
FIXt(0.188)	-0.31	-0.75	14.16	0.88	-0.20	2.20	-0.49	223.28
FIXt(0.500)	-0.47	30.51	36.01	20.05	4.78	48.30	47.02	223.37

Table 3.15: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.75$), Random Stratified Sampling (STSRS) with sample size $n = 500$.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	-0.24	-0.14	-0.79	-0.59	-0.49	0.21	0.58	2.12
CVp20	0.68	0.33	-0.72	0.07	0.32	0.31	0.69	2.57
Relative Fit	2.32	1.74	0.34	1.39	1.97	1.69	2.06	3.05
MIN$\{\hat{\sigma}^2\}$	23.52	22.63	20.20	22.04	22.94	22.50	22.77	17.80
MIN$\{\hat{\sigma}^2_{CV}\}$	-0.06	-0.21	-0.49	-0.19	-0.15	-0.03	0.10	-0.55
MIN$\{\hat{\mathrm{V}}\}$	23.52	22.63	20.20	22.04	22.94	22.50	22.77	17.80
MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.07	0.02	-0.09	-0.05	0.01	0.09	0.21	-0.38
FIXt(0.010)	23.52	22.63	20.20	22.04	22.94	22.50	22.77	17.80
FIXt(0.027)	8.03	7.25	5.12	6.88	6.59	7.15	7.36	3.03
FIXt(0.071)	1.41	0.68	-1.00	0.64	0.72	0.63	0.80	-1.07
FIXt(0.188)	-0.25	-0.88	-0.10	-1.00	-0.54	-0.76	-0.63	20.54
FIXt(0.500)	-0.59	2.31	1.91	0.82	-0.69	3.19	5.04	19.00

Table 3.16: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.25$), Random Stratified Sampling (STSRS) with sample size $n = 500$.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	-0.76	2.46	-3.76	-1.59	-0.97	0.41	2.02	-3.15
CVp20	-0.92	0.62	-3.16	-1.24	-0.52	-1.23	-0.19	-2.50
Relative Fit	4.80	2.90	-5.27	-3.02	4.12	-0.16	2.38	35.00
MIN$\{\hat{\sigma}^2\}$	31.08	24.39	13.14	17.09	37.09	15.94	23.62	3.70
MIN$\{\hat{\sigma}^2_{CV}\}$	-1.54	-1.38	-4.42	-2.15	-0.10	-1.54	-1.25	-2.64
MIN$\{\hat{\mathrm{V}}\}$	31.08	24.39	13.14	17.09	37.09	15.94	23.61	3.69
MIN$\{\hat{\mathrm{V}}_{CV}\}$	-1.09	-0.56	-2.46	-1.77	-0.33	-1.22	-0.59	-1.67
FIXt(0.010)	31.08	24.39	13.14	17.09	37.09	15.94	23.62	3.70
FIXt(0.027)	20.45	14.96	4.11	8.42	19.10	8.31	13.82	-2.01
FIXt(0.071)	8.04	3.17	-5.47	0.11	2.69	-1.85	2.28	0.32
FIXt(0.188)	-1.04	-2.92	0.32	-2.93	-2.70	-2.69	-3.01	130.85
FIXt(0.500)	-3.89	32.40	18.18	17.86	-0.26	40.25	38.62	135.58

Table 3.17: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.75$), Random Stratified Sampling (STSRS) with sample size $n = 100$.

Averaging	Population Functions							
Method	LINE	QUAD	BUMP	JUMP	NCDF	EXPO	SLOW	FAST
CV3	-0.18	-2.64	-2.17	-2.20	-0.71	-2.62	-2.87	-2.95
CVp20	-0.34	-2.41	-1.96	-2.21	-0.53	-2.48	-2.53	-2.87
Relative Fit	5.48	1.97	1.72	2.49	4.69	1.50	1.61	-4.88
MIN$\{\hat{\sigma}^2\}$	32.03	27.17	26.53	28.11	31.20	26.02	26.76	13.08
MIN$\{\hat{\sigma}^2_{CV}\}$	-1.30	-1.74	-1.82	-1.44	-1.16	-1.63	-1.80	-4.75
MIN$\{\hat{\mathrm{V}}\}$	32.03	27.17	26.53	28.11	31.20	26.02	26.76	13.08
MIN$\{\hat{\mathrm{V}}_{CV}\}$	-0.82	-1.31	-1.42	-1.31	-0.63	-1.28	-1.27	-3.56
FIXt(0.010)	32.03	27.17	26.53	28.11	31.20	26.02	26.76	13.08
FIXt(0.027)	21.35	17.01	16.31	17.80	19.87	16.14	16.54	4.23
FIXt(0.071)	8.77	4.82	4.51	5.85	6.62	4.12	4.47	-4.92
FIXt(0.188)	-0.53	-3.86	-2.17	-3.02	-0.96	-4.10	-4.01	5.84
FIXt(0.500)	-3.62	-2.10	-2.57	-3.22	-2.71	-1.33	-1.75	3.69

Table 3.18: Mean Squared Error (MSE) of 12 model averaging methods relative to method CV minus one (in percent) for eight populations ($R^2 = 0.25$), Random Stratified Sampling (STSRS) with sample size $n = 100$.

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.00	0.00	1.00
$R^2 = 0.75$		(0.00)	(0.00)	(0.01)	(0.01)	(0.02)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.04	0.08	0.11	0.73
		(0.07)	(0.09)	(0.16)	(0.22)	(0.26)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.04	0.08	0.11	0.73
		(0.07)	(0.09)	(0.16)	(0.22)	(0.26)
$n = 500$	CV	0.00	0.00	0.00	0.00	1.00
$R^2 = 0.25$		(0.00)	(0.00)	(0.01)	(0.01)	(0.02)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.04	0.08	0.09	0.76
		(0.07)	(0.09)	(0.16)	(0.20)	(0.25)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.04	0.08	0.09	0.76
		(0.07)	(0.09)	(0.16)	(0.20)	(0.25)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.75$		(0.02)	(0.02)	(0.02)	(0.09)	(0.10)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.18
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.04	0.07	0.11	0.74
		(0.08)	(0.10)	(0.15)	(0.23)	(0.27)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.04	0.07	0.11	0.74
		(0.08)	(0.09)	(0.15)	(0.23)	(0.27)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.02)	(0.02)	(0.02)	(0.09)	(0.10)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.18
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.04	0.06	0.10	0.76
		(0.08)	(0.09)	(0.15)	(0.21)	(0.26)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.04	0.06	0.09	0.76
		(0.08)	(0.09)	(0.14)	(0.21)	(0.26)

Table 3.19: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Linear' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.00	1.00	0.00
$R^2 = 0.75$		(0.01)	(0.00)	(0.01)	(0.02)	(0.02)
SRS	Relative Fit	0.20	0.19	0.19	0.19	0.23
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.56	0.33	0.00
		(0.08)	(0.13)	(0.22)	(0.14)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.26	0.64	0.00
		(0.08)	(0.11)	(0.25)	(0.23)	(0.00)
$n = 500$	CV	0.00	0.00	0.00	0.00	1.00
$R^2 = 0.25$		(0.00)	(0.01)	(0.01)	(0.03)	(0.03)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.15	0.76	0.01
		(0.08)	(0.10)	(0.22)	(0.24)	(0.02)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.04	0.13	0.74	0.05
		(0.08)	(0.10)	(0.21)	(0.26)	(0.05)
$n = 100$	CV	0.00	0.00	0.01	0.11	0.88
$R^2 = 0.75$		(0.02)	(0.03)	(0.08)	(0.31)	(0.32)
SRS	Relative Fit	0.22	0.20	0.18	0.18	0.22
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.16	0.74	0.00
		(0.08)	(0.11)	(0.23)	(0.24)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.12	0.78	0.00
		(0.08)	(0.11)	(0.21)	(0.23)	(0.01)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.02)	(0.02)	(0.03)	(0.09)	(0.10)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.10	0.55	0.27
		(0.08)	(0.10)	(0.19)	(0.29)	(0.19)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.09	0.42	0.40
		(0.08)	(0.10)	(0.18)	(0.30)	(0.24)

Table 3.20: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Quadratic' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.91	0.09	0.00
$R^2 = 0.75$		(0.01)	(0.01)	(0.28)	(0.28)	(0.00)
SRS	Relative Fit	0.20	0.19	0.19	0.20	0.22
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.22	0.72	0.00	0.00
		(0.09)	(0.19)	(0.18)	(0.00)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.22	0.72	0.00	0.00
		(0.09)	(0.19)	(0.18)	(0.00)	(0.00)
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.01)	(0.01)	(0.03)	(0.09)	(0.10)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.08	0.64	0.10	0.13
		(0.08)	(0.14)	(0.26)	(0.18)	(0.11)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.08	0.63	0.09	0.14
		(0.08)	(0.14)	(0.26)	(0.18)	(0.12)
$n = 100$	CV	0.00	0.00	0.04	0.18	0.77
$R^2 = 0.75$		(0.02)	(0.06)	(0.19)	(0.39)	(0.42)
SRS	Relative Fit	0.22	0.20	0.18	0.19	0.21
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.08	0.66	0.15	0.06
		(0.09)	(0.15)	(0.26)	(0.20)	(0.09)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.08	0.66	0.14	0.07
		(0.09)	(0.15)	(0.26)	(0.20)	(0.09)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.02)	(0.03)	(0.04)	(0.09)	(0.11)
SRS	Relative Fit	0.23	0.21	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.06	0.17	0.20	0.53
		(0.08)	(0.11)	(0.23)	(0.30)	(0.31)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.06	0.17	0.18	0.54
		(0.08)	(0.11)	(0.23)	(0.29)	(0.31)

Table 3.21: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Bump' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.02	0.18	0.77	0.04	0.00
$R^2 = 0.75$		(0.13)	(0.38)	(0.42)	(0.20)	(0.00)
SRS	Relative Fit	0.19	0.18	0.19	0.20	0.24
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.20	0.50	0.27	0.03	0.00
		(0.16)	(0.26)	(0.18)	(0.03)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.20	0.50	0.27	0.03	0.00
		(0.16)	(0.26)	(0.18)	(0.03)	(0.00)
$n = 500$	CV	0.00	0.00	0.00	0.02	0.98
$R^2 = 0.25$		(0.01)	(0.00)	(0.02)	(0.15)	(0.15)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.08	0.19	0.67	0.00
		(0.09)	(0.14)	(0.22)	(0.21)	(0.01)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.08	0.18	0.67	0.00
		(0.09)	(0.14)	(0.22)	(0.21)	(0.01)
$n = 100$	CV	0.00	0.01	0.04	0.49	0.45
$R^2 = 0.75$		(0.06)	(0.11)	(0.20)	(0.50)	(0.50)
SRS	Relative Fit	0.21	0.19	0.18	0.19	0.22
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.09	0.14	0.44	0.33	0.00
		(0.13)	(0.19)	(0.30)	(0.22)	(0.01)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.09	0.14	0.43	0.34	0.01
		(0.13)	(0.19)	(0.30)	(0.22)	(0.02)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.02)	(0.03)	(0.04)	(0.09)	(0.11)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.10	0.56	0.24
		(0.08)	(0.11)	(0.19)	(0.32)	(0.24)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.10	0.49	0.31
		(0.08)	(0.11)	(0.19)	(0.34)	(0.28)

Table 3.22: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Jump' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.75$		(0.03)	(0.03)	(0.04)	(0.07)	(0.09)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.07	0.13	0.75	0.01
		(0.08)	(0.14)	(0.24)	(0.28)	(0.03)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.07	0.13	0.72	0.04
		(0.08)	(0.14)	(0.24)	(0.30)	(0.07)
$n = 500$	CV	0.00	0.00	0.00	0.00	1.00
$R^2 = 0.25$		(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.02	0.07	0.12	0.08	0.71
		(0.06)	(0.13)	(0.20)	(0.20)	(0.26)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.02	0.07	0.12	0.08	0.72
		(0.06)	(0.12)	(0.20)	(0.20)	(0.26)
$n = 100$	CV	0.01	0.01	0.01	0.04	0.94
$R^2 = 0.75$		(0.09)	(0.09)	(0.09)	(0.19)	(0.24)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.19
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.10	0.42	0.36
		(0.11)	(0.14)	(0.21)	(0.35)	(0.29)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.10	0.33	0.45
		(0.11)	(0.14)	(0.21)	(0.34)	(0.32)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.02)	(0.03)	(0.04)	(0.08)	(0.10)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.18
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.04	0.07	0.10	0.75
		(0.08)	(0.10)	(0.16)	(0.22)	(0.27)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.04	0.07	0.09	0.75
		(0.08)	(0.10)	(0.16)	(0.22)	(0.27)

Table 3.23: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Normal CDF' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.00	1.00	0.00
$R^2 = 0.75$		(0.01)	(0.00)	(0.03)	(0.05)	(0.03)
SRS	Relative Fit	0.20	0.19	0.19	0.19	0.22
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.07	0.73	0.14	0.00
		(0.08)	(0.14)	(0.20)	(0.09)	(0.01)
	MIN$\{\hat{\text{V}}_{CV}\}$	0.05	0.07	0.67	0.20	0.01
		(0.08)	(0.14)	(0.21)	(0.12)	(0.01)
$n = 500$	CV	0.00	0.00	0.00	0.00	1.00
$R^2 = 0.25$		(0.01)	(0.00)	(0.01)	(0.03)	(0.03)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.18	0.69	0.04
		(0.08)	(0.11)	(0.24)	(0.28)	(0.06)
	MIN$\{\hat{\text{V}}_{CV}\}$	0.04	0.05	0.16	0.67	0.08
		(0.08)	(0.10)	(0.23)	(0.29)	(0.08)
$n = 100$	CV	0.00	0.00	0.01	0.12	0.87
$R^2 = 0.75$		(0.02)	(0.03)	(0.08)	(0.33)	(0.34)
SRS	Relative Fit	0.22	0.20	0.18	0.18	0.21
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.27	0.62	0.01
		(0.09)	(0.12)	(0.27)	(0.27)	(0.02)
	MIN$\{\hat{\text{V}}_{CV}\}$	0.05	0.06	0.24	0.65	0.01
		(0.09)	(0.12)	(0.26)	(0.27)	(0.03)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.02)	(0.02)	(0.02)	(0.09)	(0.10)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.10	0.38	0.44
		(0.08)	(0.10)	(0.19)	(0.32)	(0.28)
	MIN$\{\hat{\text{V}}_{CV}\}$	0.04	0.05	0.10	0.33	0.48
		(0.08)	(0.10)	(0.18)	(0.32)	(0.29)

Table 3.24: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\text{V}}_{CV}\}$ for population function 'Exponential' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.00	1.00	0.00
$R^2 = 0.75$		(0.01)	(0.01)	(0.06)	(0.07)	(0.00)
SRS	Relative Fit	0.20	0.19	0.19	0.19	0.24
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.06	0.80	0.10	0.00
		(0.08)	(0.12)	(0.18)	(0.06)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.06	0.80	0.10	0.00
		(0.08)	(0.12)	(0.18)	(0.06)	(0.00)
$n = 500$	CV	0.00	0.00	0.00	0.02	0.98
$R^2 = 0.25$		(0.00)	(0.01)	(0.01)	(0.14)	(0.14)
SRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.23	0.67	0.00
		(0.08)	(0.11)	(0.25)	(0.24)	(0.01)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.23	0.67	0.00
		(0.08)	(0.11)	(0.25)	(0.25)	(0.01)
$n = 100$	CV	0.00	0.00	0.01	0.48	0.51
$R^2 = 0.75$		(0.02)	(0.03)	(0.09)	(0.50)	(0.50)
SRS	Relative Fit	0.22	0.20	0.18	0.18	0.22
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.29	0.61	0.00
		(0.09)	(0.12)	(0.25)	(0.24)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.28	0.61	0.00
		(0.09)	(0.12)	(0.26)	(0.24)	(0.00)
$n = 100$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.02)	(0.02)	(0.03)	(0.10)	(0.11)
SRS	Relative Fit	0.23	0.21	0.19	0.18	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.11	0.55	0.26
		(0.08)	(0.10)	(0.19)	(0.32)	(0.24)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.11	0.54	0.26
		(0.08)	(0.10)	(0.19)	(0.33)	(0.25)

Table 3.25: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Slow sine' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.98	0.02	0.00	0.00
$R^2 = 0.75$		(0.02)	(0.13)	(0.13)	(0.00)	(0.00)
SRS	Relative Fit	0.15	0.15	0.16	0.27	0.27
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.93	0.02	0.00	0.00
		(0.10)	(0.11)	(0.03)	(0.00)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.93	0.02	0.00	0.00
		(0.10)	(0.11)	(0.03)	(0.00)	(0.00)
$n = 500$	CV	0.00	0.00	0.97	0.01	0.02
$R^2 = 0.25$		(0.02)	(0.02)	(0.18)	(0.09)	(0.15)
SRS	Relative Fit	0.20	0.19	0.19	0.21	0.21
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.27	0.67	0.00	0.00
		(0.10)	(0.20)	(0.17)	(0.00)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.27	0.67	0.00	0.00
		(0.10)	(0.20)	(0.17)	(0.00)	(0.00)
$n = 100$	CV	0.00	0.04	0.96	0.00	0.00
$R^2 = 0.75$		(0.07)	(0.18)	(0.20)	(0.00)	(0.00)
SRS	Relative Fit	0.18	0.16	0.16	0.25	0.26
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.08	0.36	0.56	0.00	0.00
		(0.12)	(0.21)	(0.18)	(0.00)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.08	0.35	0.57	0.00	0.00
		(0.12)	(0.21)	(0.18)	(0.00)	(0.00)
$n = 100$	CV	0.00	0.00	0.04	0.02	0.93
$R^2 = 0.25$		(0.03)	(0.06)	(0.21)	(0.13)	(0.25)
SRS	Relative Fit	0.22	0.20	0.18	0.20	0.20
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.09	0.61	0.01	0.24
		(0.10)	(0.15)	(0.27)	(0.05)	(0.19)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.09	0.61	0.01	0.24
		(0.10)	(0.15)	(0.27)	(0.05)	(0.19)

Table 3.26: Model Averaging coefficients and corresponding standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Fast sine' and Simple Random Sampling (SRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.75$		(0.03)	(0.03)	(0.03)	(0.07)	(0.09)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.07	0.13	0.70
		(0.08)	(0.11)	(0.16)	(0.24)	(0.28)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.05	0.06	0.13	0.71
		(0.09)	(0.12)	(0.16)	(0.26)	(0.30)
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.03)	(0.03)	(0.02)	(0.07)	(0.08)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.07	0.09	0.75
		(0.08)	(0.10)	(0.16)	(0.21)	(0.26)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.05	0.06	0.10	0.74
		(0.09)	(0.12)	(0.16)	(0.24)	(0.29)
$n = 100$	CV	0.00	0.01	0.01	0.05	0.93
$R^2 = 0.75$		(0.05)	(0.08)	(0.10)	(0.22)	(0.25)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.18
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.04	0.08	0.14	0.70
		(0.08)	(0.10)	(0.17)	(0.26)	(0.30)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.08	0.13	0.70
		(0.09)	(0.12)	(0.18)	(0.26)	(0.32)
$n = 100$	CV	0.00	0.01	0.01	0.05	0.93
$R^2 = 0.25$		(0.05)	(0.08)	(0.10)	(0.22)	(0.25)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.18
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.04	0.07	0.11	0.73
		(0.08)	(0.10)	(0.16)	(0.24)	(0.29)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.08	0.11	0.72
		(0.09)	(0.12)	(0.17)	(0.25)	(0.31)

Table 3.27: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Linear' and Random Stratified Sampling (STSRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.01	0.70	0.29
$R^2 = 0.75$		(0.03)	(0.03)	(0.07)	(0.46)	(0.45)
STSRS	Relative Fit	0.20	0.19	0.19	0.19	0.22
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.07	0.45	0.42	0.00
		(0.09)	(0.14)	(0.25)	(0.19)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.07	0.19	0.68	0.00
		(0.10)	(0.14)	(0.26)	(0.26)	(0.01)
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.03)	(0.03)	(0.03)	(0.08)	(0.10)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.05	0.12	0.75	0.02
		(0.08)	(0.12)	(0.21)	(0.25)	(0.04)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.10	0.68	0.12
		(0.09)	(0.13)	(0.21)	(0.29)	(0.13)
$n = 100$	CV	0.00	0.01	0.04	0.27	0.68
$R^2 = 0.75$		(0.06)	(0.11)	(0.19)	(0.44)	(0.47)
STSRS	Relative Fit	0.22	0.20	0.18	0.18	0.21
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.06	0.15	0.74	0.00
		(0.09)	(0.12)	(0.24)	(0.26)	(0.02)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.14	0.73	0.02
		(0.10)	(0.14)	(0.24)	(0.29)	(0.06)
$n = 100$	CV	0.00	0.01	0.01	0.05	0.93
$R^2 = 0.25$		(0.05)	(0.09)	(0.10)	(0.22)	(0.26)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.10	0.43	0.39
		(0.08)	(0.11)	(0.19)	(0.33)	(0.28)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.10	0.32	0.49
		(0.10)	(0.13)	(0.20)	(0.34)	(0.34)

Table 3.28: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Quadratic' and Random Stratified Sampling (STSRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.01	0.21	0.61	0.17
$R^2 = 0.75$		(0.05)	(0.08)	(0.41)	(0.49)	(0.38)
STSRS	Relative Fit	0.20	0.19	0.19	0.20	0.22
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.15	0.79	0.00	0.00
		(0.10)	(0.19)	(0.20)	(0.01)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.14	0.77	0.02	0.01
		(0.11)	(0.21)	(0.24)	(0.05)	(0.02)
$n = 500$	CV	0.00	0.00	0.00	0.01	0.98
$R^2 = 0.25$		(0.04)	(0.04)	(0.05)	(0.10)	(0.12)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.09	0.47	0.21	0.18
		(0.09)	(0.16)	(0.29)	(0.26)	(0.15)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.09	0.33	0.28	0.24
		(0.10)	(0.17)	(0.31)	(0.31)	(0.24)
$n = 100$	CV	0.01	0.02	0.09	0.20	0.68
$R^2 = 0.75$		(0.09)	(0.15)	(0.29)	(0.40)	(0.47)
STSRS	Relative Fit	0.22	0.20	0.18	0.19	0.21
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.10	0.49	0.28	0.09
		(0.10)	(0.16)	(0.31)	(0.28)	(0.13)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.10	0.38	0.31	0.15
		(0.12)	(0.18)	(0.33)	(0.31)	(0.20)
$n = 100$	CV	0.00	0.01	0.01	0.05	0.92
$R^2 = 0.25$		(0.06)	(0.09)	(0.12)	(0.22)	(0.27)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.06	0.14	0.21	0.55
		(0.09)	(0.12)	(0.22)	(0.31)	(0.32)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.13	0.19	0.58
		(0.10)	(0.14)	(0.22)	(0.31)	(0.34)

Table 3.29: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Bump' and Random Stratified Sampling (STSRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.01	0.02	0.57	0.40
$R^2 = 0.75$		(0.06)	(0.07)	(0.13)	(0.49)	(0.49)
STSRS	Relative Fit	0.19	0.18	0.19	0.20	0.24
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.17	0.44	0.33	0.06	0.00
		(0.16)	(0.26)	(0.21)	(0.05)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.09	0.20	0.41	0.29	0.01
		(0.13)	(0.22)	(0.30)	(0.16)	(0.02)
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.03)	(0.03)	(0.03)	(0.07)	(0.09)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.07	0.11	0.75	0.01
		(0.09)	(0.13)	(0.19)	(0.23)	(0.03)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.09	0.59	0.21
		(0.09)	(0.13)	(0.19)	(0.30)	(0.21)
$n = 100$	CV	0.01	0.02	0.05	0.19	0.73
$R^2 = 0.75$		(0.09)	(0.14)	(0.21)	(0.39)	(0.44)
STSRS	Relative Fit	0.21	0.19	0.18	0.19	0.23
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.12	0.14	0.28	0.46	0.01
		(0.14)	(0.19)	(0.28)	(0.25)	(0.03)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.08	0.09	0.18	0.57	0.08
		(0.13)	(0.16)	(0.26)	(0.31)	(0.12)
$n = 100$	CV	0.00	0.01	0.01	0.05	0.93
$R^2 = 0.25$		(0.06)	(0.08)	(0.11)	(0.22)	(0.26)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.05	0.10	0.43	0.37
		(0.09)	(0.12)	(0.19)	(0.34)	(0.29)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.05	0.09	0.26	0.55
		(0.10)	(0.13)	(0.19)	(0.33)	(0.34)

Table 3.30: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Jump' and Random Stratified Sampling (STSRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.00	0.10	0.89
$R^2 = 0.75$		(0.02)	(0.02)	(0.04)	(0.31)	(0.31)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.02	0.07	0.15	0.75	0.00
		(0.06)	(0.14)	(0.25)	(0.27)	(0.02)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.02	0.08	0.18	0.71	0.01
		(0.07)	(0.15)	(0.27)	(0.28)	(0.04)
$n = 500$	CV	0.00	0.00	0.00	0.00	0.99
$R^2 = 0.25$		(0.04)	(0.02)	(0.03)	(0.07)	(0.09)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.03	0.07	0.11	0.11	0.69
		(0.06)	(0.13)	(0.19)	(0.23)	(0.28)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.06	0.09	0.12	0.70
		(0.08)	(0.13)	(0.19)	(0.26)	(0.32)
$n = 100$	CV	0.01	0.01	0.01	0.02	0.97
$R^2 = 0.75$		(0.08)	(0.07)	(0.07)	(0.13)	(0.18)
STSRS	Relative Fit	0.23	0.20	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.10	0.48	0.30
		(0.10)	(0.14)	(0.21)	(0.35)	(0.27)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.11	0.54	0.23
		(0.10)	(0.14)	(0.22)	(0.37)	(0.31)
$n = 100$	CV	0.00	0.01	0.01	0.04	0.94
$R^2 = 0.25$		(0.06)	(0.10)	(0.11)	(0.19)	(0.24)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.08	0.12	0.72
		(0.08)	(0.11)	(0.17)	(0.25)	(0.30)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.08	0.12	0.71
		(0.09)	(0.14)	(0.18)	(0.27)	(0.33)

Table 3.31: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Normal CDF' and Random Stratified Sampling (STSRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.01	0.96	0.02
$R^2 = 0.75$		(0.04)	(0.05)	(0.10)	(0.18)	(0.14)
STSRS	Relative Fit	0.20	0.20	0.19	0.19	0.21
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.09	0.66	0.19	0.00
		(0.09)	(0.16)	(0.24)	(0.13)	(0.01)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.10	0.56	0.27	0.01
		(0.10)	(0.18)	(0.28)	(0.19)	(0.01)
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.03)	(0.03)	(0.04)	(0.10)	(0.11)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.15	0.68	0.07
		(0.08)	(0.12)	(0.23)	(0.29)	(0.09)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.12	0.67	0.09
		(0.09)	(0.13)	(0.23)	(0.30)	(0.11)
$n = 100$	CV	0.01	0.02	0.06	0.46	0.45
$R^2 = 0.75$		(0.08)	(0.13)	(0.24)	(0.50)	(0.50)
STSRS	Relative Fit	0.22	0.20	0.19	0.18	0.20
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.25	0.62	0.02
		(0.09)	(0.13)	(0.29)	(0.30)	(0.07)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.07	0.25	0.60	0.02
		(0.11)	(0.15)	(0.31)	(0.33)	(0.07)
$n = 100$	CV	0.00	0.01	0.01	0.06	0.92
$R^2 = 0.25$		(0.06)	(0.09)	(0.11)	(0.23)	(0.27)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.10	0.35	0.46
		(0.08)	(0.11)	(0.19)	(0.35)	(0.33)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.10	0.35	0.45
		(0.10)	(0.13)	(0.21)	(0.36)	(0.34)

Table 3.32: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Exponential' and Random Stratified Sampling (STSRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.00	0.00	0.01	0.95	0.04
$R^2 = 0.75$		(0.04)	(0.04)	(0.07)	(0.21)	(0.19)
STSRS	Relative Fit	0.20	0.19	0.19	0.19	0.23
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.07	0.71	0.17	0.00
		(0.09)	(0.15)	(0.21)	(0.09)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.08	0.28	0.59	0.00
		(0.10)	(0.15)	(0.27)	(0.24)	(0.00)
$n = 500$	CV	0.00	0.00	0.00	0.01	0.99
$R^2 = 0.25$		(0.03)	(0.03)	(0.04)	(0.10)	(0.11)
STSRS	Relative Fit	0.21	0.20	0.20	0.20	0.20
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.17	0.71	0.01
		(0.08)	(0.12)	(0.23)	(0.24)	(0.03)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.09	0.73	0.06
		(0.09)	(0.13)	(0.20)	(0.26)	(0.08)
$n = 100$	CV	0.00	0.01	0.04	0.37	0.58
$R^2 = 0.75$		(0.06)	(0.11)	(0.20)	(0.48)	(0.49)
STSRS	Relative Fit	0.22	0.20	0.18	0.18	0.22
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.05	0.06	0.24	0.65	0.00
		(0.09)	(0.13)	(0.27)	(0.27)	(0.01)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.05	0.06	0.16	0.72	0.02
		(0.10)	(0.14)	(0.25)	(0.29)	(0.05)
$n = 100$	CV	0.00	0.01	0.01	0.05	0.92
$R^2 = 0.25$		(0.06)	(0.08)	(0.11)	(0.22)	(0.26)
STSRS	Relative Fit	0.23	0.21	0.19	0.19	0.19
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.04	0.05	0.11	0.47	0.33
		(0.08)	(0.11)	(0.20)	(0.35)	(0.30)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.04	0.05	0.10	0.36	0.44
		(0.10)	(0.13)	(0.20)	(0.35)	(0.34)

Table 3.33: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Slow sine' and Random Stratified Sampling (STSRS).

Sampling Design	Averaging Method	Model Averaging Coefficients (Standard Errors)				
		α_1	α_2	α_3	α_4	α_5
$n = 500$	CV	0.01	0.40	0.59	0.00	0.00
$R^2 = 0.75$		(0.09)	(0.49)	(0.49)	(0.00)	(0.00)
STSRS	Relative Fit	0.16	0.15	0.16	0.26	0.27
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.86	0.08	0.00	0.00
		(0.11)	(0.14)	(0.05)	(0.00)	(0.00)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.07	0.80	0.13	0.00	0.00
		(0.13)	(0.16)	(0.08)	(0.00)	(0.00)
$n = 500$	CV	0.00	0.01	0.41	0.06	0.52
$R^2 = 0.25$		(0.05)	(0.08)	(0.49)	(0.23)	(0.50)
STSRS	Relative Fit	0.20	0.19	0.19	0.21	0.21
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.20	0.74	0.00	0.00
		(0.10)	(0.20)	(0.20)	(0.00)	(0.01)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.07	0.18	0.74	0.00	0.01
		(0.11)	(0.22)	(0.24)	(0.00)	(0.04)
$n = 100$	CV	0.02	0.13	0.84	0.00	0.00
$R^2 = 0.75$		(0.15)	(0.34)	(0.37)	(0.06)	(0.05)
STSRS	Relative Fit	0.18	0.16	0.16	0.25	0.26
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.09	0.32	0.59	0.00	0.00
		(0.13)	(0.24)	(0.22)	(0.00)	(0.02)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.11	0.28	0.61	0.00	0.01
		(0.16)	(0.27)	(0.28)	(0.01)	(0.04)
$n = 100$	CV	0.01	0.02	0.10	0.07	0.80
$R^2 = 0.25$		(0.09)	(0.15)	(0.30)	(0.26)	(0.40)
STSRS	Relative Fit	0.22	0.20	0.19	0.20	0.20
		(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
	MIN$\{\hat{\sigma}^2_{CV}\}$	0.06	0.10	0.48	0.03	0.33
		(0.10)	(0.17)	(0.31)	(0.11)	(0.23)
	MIN$\{\hat{\mathrm{V}}_{CV}\}$	0.06	0.11	0.37	0.04	0.41
		(0.12)	(0.19)	(0.32)	(0.14)	(0.28)

Table 3.34: Model Averaging coefficients and their standard errors using method CV, Relative Fit, MIN$\{\hat{\sigma}^2_{CV}\}$, and MIN$\{\hat{\mathrm{V}}_{CV}\}$ for population function 'Fast sine' and Random Stratified Sampling (STSRS).

APPENDIX A Theorems, Corollaries and Lemmas for Chapter 1

Theorem A.1. *Suppose that* $\mathbf{x}$ *is an* $n \times 1$ *random vector with mean* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$ *and* $\mathbf{A}$ *is an* $n \times n$ *symmetric matrix. Then*

$$E(\mathbf{x}^T\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$$

Proof. Note that

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = (\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu}) + \boldsymbol{\mu}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{A}\boldsymbol{\mu} - \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}. \tag{A.1}$$

Because $(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})$ is a scalar, so $(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu}) = \mathrm{tr}[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})]$. Thus

$$\begin{aligned}
\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}) &= \mathrm{E}\left[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})\right] + 2\mathrm{E}\left(\mathbf{x}^T\mathbf{A}\boldsymbol{\mu}\right) - \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{E}\left\{\mathrm{tr}[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})]\right\} + 2\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} - \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{E}\left\{\mathrm{tr}[\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T]\right\} + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\{\mathbf{A}\mathrm{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T]\} + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

□

Lemma A.1. *Suppose that* $\mathbf{x}$ *is a random vector. Its elements* x_i*,* $i = 1, ..., n$ *are mutually independent with mean* $E(x_i) = \mu_i$ *and variance* $Var(x_i) = \sigma^2\omega_i$*. Denote*

x_i's third and fourth moment $m_{ir} = E(x_i - \mu_i)^r$ and denote $\mathbf{m}_r = diag\{m_{1r}, ..., m_{nr}\}$, $r = 3,\ 4$. Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ symmetric matrix. Then

$$Var(\mathbf{x}^T\mathbf{A}\mathbf{x}) = (\mathbf{a}^T\mathbf{m}_4\mathbf{a} - 3\sigma^4\mathbf{a}^T\boldsymbol{\Omega}^2\mathbf{a}) + 2\sigma^4 tr(\mathbf{A}\boldsymbol{\Omega})^2 + 4\sigma^2\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\Omega}\mathbf{A}\boldsymbol{\mu} + 4\boldsymbol{\mu}^T\mathbf{A}\mathbf{m}_3\mathbf{a}$$

where $\mathbf{a}^T = (a_{11}, ..., a_{nn})$, $\boldsymbol{\Omega} = diag\{\omega_1, \omega_2, \cdots, \omega_n\}$ and $\boldsymbol{\mu}^T = (\mu_1, ..., \mu_n)$.

Proof. Note that

$$\mathrm{Var}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \mathrm{E}[(\mathbf{x}^T\mathbf{A}\mathbf{x})^2] - [\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x})]^2. \tag{A.2}$$

By Theorem A.1,

$$\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}. \tag{A.3}$$

So it remains to calculate $\mathrm{E}[(\mathbf{x}^T\mathbf{A}\mathbf{x})^2]$. To proceed, note that by (A.1),

$$\begin{aligned}(\mathbf{x}^T\mathbf{A}\mathbf{x})^2 \quad &= \quad \left[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})\right]^2 + 4\left[\boldsymbol{\mu}^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})\right]^2 + (\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu})^2 \\ &\quad +2\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}\left[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu}) + 2\boldsymbol{\mu}^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})\right] \\ &\quad +4\boldsymbol{\mu}^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu}).\end{aligned}$$

Let $\mathbf{z} = \mathbf{x} - \boldsymbol{\mu}$, then $\mathrm{E}(\mathbf{z}) = \mathbf{0}$. From Theorem A.1,

$$\begin{aligned}\mathrm{E}[(\mathbf{x}^T\mathbf{A}\mathbf{x})^2] \quad &= \quad \mathrm{E}[(\mathbf{z}^T\mathbf{A}\mathbf{z})^2] + 4\mathrm{E}[(\boldsymbol{\mu}^T\mathbf{A}\mathbf{z})^2] + (\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu})^2 \\ &\quad +2\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}[\mathrm{tr}(\mathbf{A}\boldsymbol{\Omega})]\sigma^2 + 4\mathrm{E}(\boldsymbol{\mu}^T\mathbf{A}\mathbf{z}\mathbf{z}^T\mathbf{A}\mathbf{z}).\end{aligned} \tag{A.4}$$

Now let us calculate each term on the right hand side of (A.4). For the first term, note that

$$(\mathbf{z}^T\mathbf{A}\mathbf{z})^2 = \sum_i\sum_j\sum_k\sum_l a_{ij}a_{kl}z_iz_jz_kz_l,$$

where z_i's are independent. So

$$\mathrm{E}(z_iz_jz_kz_l) = \begin{cases} m_{4i}, & \text{if} \quad i=j=k=l \\ \sigma^4\omega_i\omega_k, & \text{if} \quad i=j, k=l \\ \sigma^4\omega_i\omega_j, & \text{if} \quad i=k, j=l;\ i=l, j=k \\ 0 & \text{Otherwise} \end{cases}.$$

Thus

$$\begin{aligned}
\mathrm{E}[(\mathbf{z}^T\mathbf{A}\mathbf{z})^2] &= \sum_i a_{ii}^2 m_{4i} + \sigma^4 \sum \sum_{i \neq k} a_{ii}a_{kk}\omega_i\omega_k + \sigma^4 \sum \sum_{i \neq j} a_{ij}^2 \omega_i\omega_j \\
&\quad + \sigma^4 \sum \sum_{i \neq j} a_{ij}a_{ji}\omega_i\omega_j \\
&= \sum_i a_{ii}^2 m_{4i} + \sigma^4 \sum_i \sum_k a_{ii}a_{kk}\omega_i\omega_k + \sigma^4 \sum_i \sum_j a_{ij}^2 \omega_i\omega_j \\
&\quad + \sigma^4 \sum_i \sum_j a_{ij}a_{ji}\omega_i\omega_j - 3\sigma^4 \sum_i a_i^2\omega_i^2 \\
&= (\mathbf{a}^T\mathbf{m}_4\mathbf{a} - 3\sigma^4\mathbf{a}^T\boldsymbol{\Omega}^2\mathbf{a}) + \sigma^4[\mathrm{tr}(\mathbf{A}\boldsymbol{\Omega})]^2 + 2\sigma^4\mathrm{tr}[(\mathbf{A}\boldsymbol{\Omega})^2]. \qquad (A.5)
\end{aligned}$$

For the second term,

$$\begin{aligned}
\mathrm{E}[(\boldsymbol{\mu}^T\mathbf{A}\mathbf{z})^2] &= \mathrm{E}(\boldsymbol{\mu}^T\mathbf{A}\mathbf{z}\boldsymbol{\mu}^T\mathbf{A}\mathbf{z}) = \mathrm{E}(\mathbf{z}^T\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}^T\mathbf{A}\mathbf{z}) \\
&= \sigma^2\mathrm{tr}(\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\Omega}) = \sigma^2\mathrm{tr}(\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\Omega}\mathbf{A}\boldsymbol{\mu}) \\
&= \sigma^2\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\Omega}\mathbf{A}\boldsymbol{\mu}. \qquad (A.6)
\end{aligned}$$

Finally, let $\mathbf{c} = \mathbf{A}\boldsymbol{\mu}$, then

$$\mathrm{E}(\boldsymbol{\mu}^T\mathbf{A}\mathbf{z}\mathbf{z}^T\mathbf{A}\mathbf{z}) = \sum_i \sum_j \sum_k c_i a_{jk} \mathrm{E}(z_i z_j z_k)$$

Note that

$$\mathrm{E}(z_i z_j z_k) = \begin{cases} m_{3i}, & \text{if} \quad i = j = k \\ 0 & \text{Otherwise} \end{cases},$$

Therefore

$$\mathrm{E}(\boldsymbol{\mu}^T\mathbf{A}\mathbf{z}\mathbf{z}^T\mathbf{A}\mathbf{z}) = \sum_i c_i a_{ii} m_{3i} = \mathbf{c}^T\mathbf{m}_3\mathbf{a} = \boldsymbol{\mu}^T\mathbf{A}\mathbf{m}_3\mathbf{a}. \qquad (A.7)$$

The result follows by substituting (A.5), (A.6) and (A.7) into (A.4), and substituting (A.3) and (A.4) into (A.2). □

Chebyshev's inequality is one of the most important tools for establishing the order in probability of random variables.

Theorem A.2. *Let X be a random variable such that $E(|X|^r) < \infty$, where $r > 0$. Let $F(x)$ be the distribution function of X. Then, for every $\varepsilon > 0$ and finite A,*

$$P(|X - A| \geq \varepsilon) \leq \frac{E(|X - A|^r)}{\varepsilon^r}.$$

Proof. Let S denote the set of x for which $|x - A| \geq \varepsilon$ and let S^c denote the set of x for which $|x - A| < \varepsilon$. Then,

$$\begin{aligned} \int |x - A|^r dF(x) &= \int_S |x - A|^r dF(x) + \int_{S^c} |x - A|^r dF(x) \\ &\geq \varepsilon^r \int_S dF(x) = \varepsilon^r \mathrm{P}(|X - A| \geq \varepsilon). \end{aligned}$$

□

Corollary A.1. *Let $\{X_n\}$ be a sequence of random variables. Suppose $\{a_n\}$ is a sequence of positive real numbers such that*

$$EX_n^2 = O(a_n^2). \tag{A.8}$$

Then

$$X_n = O_p(a_n).$$

Proof. Equation (A.8) implies that there exists a finite number M such that

$$\mathrm{E}X_n^2 \leq M a_n^2,$$

for all n. By Theorem A.2, for any $\varepsilon > 0$,

$$\mathrm{P}(|X_n| \geq \varepsilon a_n) \leq \frac{\mathrm{E}X_n^2}{\varepsilon^2 a_n^2} \leq \frac{M a_n^2}{\varepsilon^2 a_n^2} = \frac{M}{\varepsilon^2},$$

and thus the result follows. □

Corollary A.2. *Let $\{X_n\}$ be a sequence of random variables. Suppose $\{a_n\}$ is a sequence of positive real numbers such that*

$$E|X_n| = O(a_n). \tag{A.9}$$

Then

$$X_n = O_p(a_n).$$

Proof. Equation (A.9) implies that there exists a finite number M_1 such that

$$\mathrm{E}|X_n| \leq M_1 a_n,$$

for all n. By Theorem A.2,

$$\mathrm{P}(|X_n| \geq M_2 a_n) \leq \frac{E|X_n|}{M_2 a_n}.$$

Hence, given $\varepsilon > 0$, we choose $M_2 \geq M_1 \varepsilon^{-1}$. Then

$$\mathrm{P}(|X_n| \geq M_2 a_n) \leq \frac{\varepsilon \mathrm{E}|X_n|}{M_1 a_n} \leq \varepsilon,$$

and the result follows. □

APPENDIX B Using R to calculate α_k's in (3.10) to (3.13) in Chapter 3

In order to choose α_k's to minimize (3.10) to (3.13), we use the *solve.QP* function in R. It chooses $\mathbf{b}$ to solve quadratic programming problems of the form $\min\{-\mathbf{d}^T\mathbf{b} + \frac{1}{2}\mathbf{b}^T\mathbf{D}\mathbf{b}\}$ with the constraints $\mathbf{A}^T\mathbf{b} >= \mathbf{b}_0$. We will divide the discussion into two sections, one for the equal selection probability SRS design and the other for the unequal selection probability STSRS design that we have described before.

Simple random sampling (SRS)

The R code can be simplified for this equal selection probability design. Choosing α_k's to minimize (3.10) is equivalent to choosing α_k's to solve

$$\begin{aligned} & \min\{(\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha})^T(\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha})\} \\ = \ & \min\{-2\mathbf{Y}^T\hat{\mathbf{m}}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\hat{\mathbf{m}}^T\hat{\mathbf{m}}\boldsymbol{\alpha}\}. \end{aligned}$$

Therefore, we can let $\mathbf{D} = 2\hat{\mathbf{m}}^T\hat{\mathbf{m}}$ and let $\mathbf{d} = 2\hat{\mathbf{m}}^T\mathbf{Y}$ to use *solve.QP* for (3.10).

To minimize (3.12), we carry out the following procedures. Note that

$$\begin{aligned} \hat{V}(t^{MA}) &= \sum\sum_s \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}} \frac{Y_i - \hat{m}_i^{MA}}{\pi_i} \frac{Y_j - \hat{m}_j^{MA}}{\pi_j} \\ &= N^2(1-f)s^2/n, \end{aligned}$$

where $s^2 = \frac{1}{n-1}\sum_s(Y_i - \hat{m}_i^{MA} - (\bar{Y}_s - \bar{\hat{m}}_s^{MA}))^2$, so

$$\begin{aligned}
\min\{\hat{V}(t^{MA})\} &= \min\{\sum_s (Y_i - \hat{m}_i^{MA} - (\bar{Y}_s - \bar{\hat{m}}_s^{MA}))^2\} \\
&= \min\{\sum_s (Y_i - \hat{m}_i^{MA})^2 - \frac{1}{n}\left[\sum_s (Y_i - \hat{m}_i^{MA})\right]^2\} \\
&= \min\{(\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha})^T(\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha}) - (\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha})^T\frac{\mathbf{1}\mathbf{1}^T}{n}(\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha})\} \\
&= \min\{(\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha})^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)(\mathbf{Y} - \hat{\mathbf{m}}\boldsymbol{\alpha})\} \\
&= \min\{-2\mathbf{Y}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\hat{\mathbf{m}}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}\boldsymbol{\alpha}\}
\end{aligned}$$

Let $\mathbf{D} = 2\hat{\mathbf{m}}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}$ and let $\mathbf{d} = 2\hat{\mathbf{m}}^T(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{Y}$ to use *solve.QP*.

To use *solve.QP* for (3.11) and (3.13), we can simply replace $\hat{\mathbf{m}}$ by $\hat{\mathbf{m}}^{(-)}$ in the above discussion for (3.10) and (3.12), respectively.

Random stratified sampling (STSRS)

For STSRS design, minimizing (3.10) is equivalent to

$$\begin{aligned}
& \min\{\sum_{h=1}^{H}\frac{1}{\pi_h}(\mathbf{Y}_h - \hat{\mathbf{m}}_h\boldsymbol{\alpha})^T(\mathbf{Y}_h - \hat{\mathbf{m}}_h\boldsymbol{\alpha})\} \\
= \; & \min\{-2\sum_{h=1}^{H}\frac{1}{\pi_h}\mathbf{Y}_h^T\hat{\mathbf{m}}_h\boldsymbol{\alpha} + \sum_{h=1}^{H}\frac{1}{\pi_h}\boldsymbol{\alpha}^T\hat{\mathbf{m}}_h^T\hat{\mathbf{m}}_h\boldsymbol{\alpha}\},
\end{aligned}$$

where $\pi_h = n_h/N_h$. So let $\mathbf{D} = 2\sum_{h=1}^{H}\frac{1}{\pi_h}\hat{\mathbf{m}}_h^T\hat{\mathbf{m}}_h$ and $\mathbf{d} = 2(\sum_{h=1}^{H}\frac{1}{\pi_h}\mathbf{Y}_h^T\hat{\mathbf{m}}_h)^T$ for *solve.QP*.

To minimize (3.12), note that

$$\begin{aligned}
\hat{V}(t^{MA}) &= \sum\sum_s \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}}\frac{Y_i - \hat{m}_i^{MA}}{\pi_i}\frac{Y_j - \hat{m}_j^{MA}}{\pi_j} \\
&= \sum_{h=1}^{H} N_h^2(1 - f_h)s_h^2/n_h
\end{aligned}$$

where $s_h^2 = \frac{1}{n_h - 1}\sum_s (Y_{hi} - \hat{m}_{hi}^{MA} - (\bar{Y}_{hs} - \bar{\hat{m}}_{hs}^{MA}))^2$, so

$$\min\{\sum_{h=1}^{H} N_h^2\frac{(1 - f_h)}{n_h(n_h - 1)}\sum_s (Y_{hi} - \hat{m}_{hi}^{MA} - (\bar{Y}_{hs} - \bar{\hat{m}}_{hs}^{MA}))^2\}$$

$$
\begin{aligned}
&\equiv \min\{\sum_{h=1}^{H} C_h(\mathbf{Y}_h - \hat{\mathbf{m}}_h\boldsymbol{\alpha})^T(\mathbf{I}_{n_h} - \frac{1}{n_h}\mathbf{1}\mathbf{1}^T)(\mathbf{Y}_h - \hat{\mathbf{m}}_h\boldsymbol{\alpha})\} \\
&= \min\{\sum_{h=1}^{H} C_h\left[-2\mathbf{Y}_h^T(\mathbf{I}_{n_h} - \frac{1}{n_h}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}_h\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\hat{\mathbf{m}}_h^T(\mathbf{I}_{n_h} - \frac{1}{n_h}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}_h\boldsymbol{\alpha}\right]\} \\
&= \min\{-2\left(\sum_{h=1}^{H} C_h\mathbf{Y}_h^T(\mathbf{I}_{n_h} - \frac{1}{n_h}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}_h\right)\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\sum_{h=1}^{H} C_h\hat{\mathbf{m}}_h^T(\mathbf{I}_{n_h} - \frac{1}{n_h}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}_h\boldsymbol{\alpha}\}
\end{aligned}
$$

where $C_h = N_h^2\frac{(1-f_h)}{n_h(n_h-1)}$. So let

$$\mathbf{D} = 2\sum_{h=1}^{H} C_h\hat{\mathbf{m}}_h^T(\mathbf{I}_{n_h} - \frac{1}{n_h}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}_h$$

and

$$\mathbf{d} = 2\left(\sum_{h=1}^{H} C_h\mathbf{Y}_h^T(\mathbf{I}_{n_h} - \frac{1}{n_h}\mathbf{1}\mathbf{1}^T)\hat{\mathbf{m}}_h\right)^T.$$

Then we are ready to use *solve.QP*.

To use *solve.QP* for (3.11) and (3.13), we can simply replace $\hat{\mathbf{m}}$ by $\hat{\mathbf{m}}^{(-)}$ in the above discussion for (3.10) and (3.12), respectively.

BIBLIOGRAPHY

Bailey, R. G., Avers, P. E., King, T., and McNab, W. H. (Eds.) (1994). *Ecoregions and Subregions of the United States (map)*. Washington, DC: U. S. Geological Survey. Scale 1:7,500,000, colored, accompanied by a supplementary table of map unit descriptions, prepared for the U. S. Department of Agriculture, Forest Service.

Bartolucci, F. and Montanari, G. (2006). A new class of unbiased estimators of the variance of the systematic sample mean. *Journal of Statistical Planning and Inference*, 136(4):1512–1525.

Breidt, F. J. and Opsomer, J. D. (2000). Local polynomial regresssion estimators in survey sampling. *The Annals of Statistics*, 28(4):1026–1053.

Breiman, L. (1996). Stacked regressions. *Machine Learning*, 24:49 – 64.

Frayer, W. and Furnival, G. (1999). Forest survey sampling designs: A history. *Journal of Forestry*, 97:4–8.

Gillespie, A. J. R. (1999). Rationale for a national annual forest inventory program. *Journal of Forestry*, 97:16–20.

Iachan, R. (1982). Systematic sampling: A critical review. *International Statistical Review*, 50:293–303.

Kish, L. (1965). *Survey sampling.* John Wiley & Sons, Inc.

LeBlanc, M. and Tibshirani, R. (1996). Combining estimates in regression and classification. *Journal of the American Statistical Association*, 91:1641–1650.

Madow, W. G. and Madow, L. G. (1944). On the theory of systematic sampling, I. *Annals of Mathematical Statistics*, 15:1–24.

Mahalanobis, P. C. (1946). Recent experiments in statistical sampling in the indian statistical institute. *Journal of the Royal Statistical Society*, 109:325–370.

Montanari, G. and Bartolucci, F. (1998). On estimating the variance of the systematic sample mean. *Journal of the Italian Statistical Society*, 7:185–196.

Opsomer, J. and Miller, C. (2005). Selecting the amount of smoothing in nonparametric regression estimation for complex surveys. *Nonparametric Statistics*, 17:593 – 611.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22:1346–1370.

Särndal, C.-E., Swensson, B., and Wretman, J. H. (1992). *Model assisted survey sampling*. Springer-Verlag Inc.

Törnqvist, L. (1963). The theory of replicated systematic cluster sampling with random start. *Revue de l'Institut International de Statistique*, 31:11–23.

U. S. Department of Agriculture Forest Service (1992). Forest service resource inventories: An overview. Technical report, Washington, DC.

Wand, M.-P. and Jones, M.-C. (1995). *Kernel smoothing*. Chapman & Hall Ltd.

Wolter, K. M. (1985). *Introduction to variance estimation*. Springer-Verlag Inc.

Yang, Y. (2001). Adaptive regression by mixing. *Journal of the American Statistical Association*, 96(454):574–588.

Zinger, A. (1980). Variance estimation in partially systematic sampling. *Journal of the American Statistical Association*, 75:206–211.